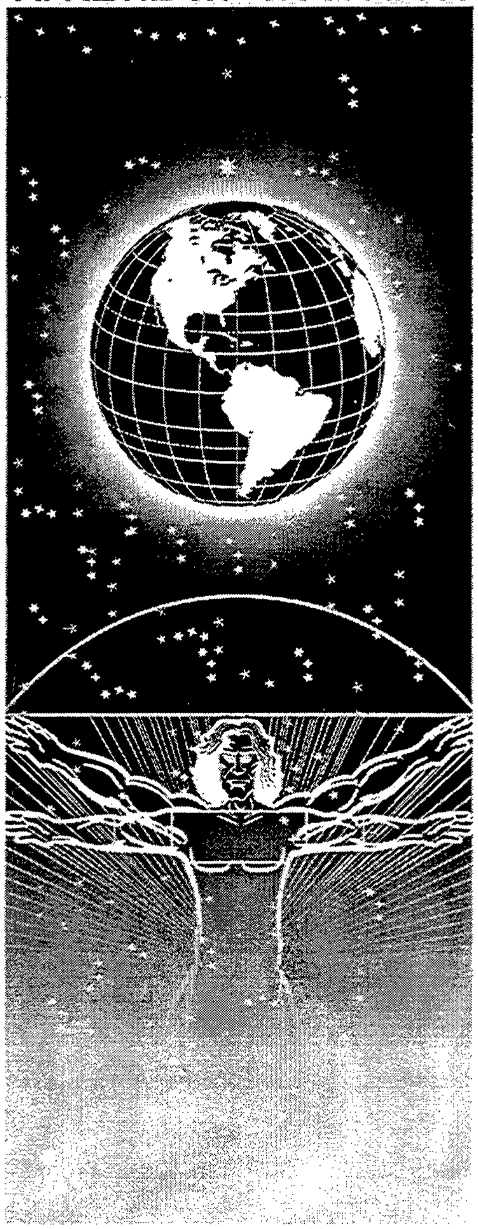
AFRL/HE-WP-TR-1999-0003



UNITED STATES AIR FORCE RESEARCH LABORATORY

AIDED AND UNAIDED OPERATOR PERFORMANCE WITH FIRST GENERATION FLIR IMAGERY

William P. Janson
Judi E. See
Joseph T. Riegler
Iris Davis

LOGICON TECHNICAL SERVICES, INC.
PO BOX 317258
DAYTON, OH 45437-7258

Gilbert G. Kuperman

HUMAN EFFECTIVENESS DIRECTORATE
CREW SYSTEM INTERFACE DIVISION
WRIGHT-PATTERSON AFB, OH 45433-7022

AUGUST 1998

INTERIM REPORT FOR THE PERIOD 1 JULY 1997 to 1 JULY 1998

Human Effectiveness Directorate
Crew System Interface Division
2255 H Street
Wright-Patterson AFB OH 45433-7022

NOTICES

TECHNICAL REVIEW AND APPROVAL

AFRL-HE-WP-TR-1999-0003

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Hendrick W Ruck

HENDRICK W. RUCK, PhD
Chief, Crew System Interface Division
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 1998	Interim Report; 1 July 1997 to 1 July 1998

4. TITLE AND SUBTITLE
Aided and Unaided Operator Performance with First Generation FLIR Imagery

5. FUNDING NUMBERS
C: F41624-94-D-6000
P: 62202F
PR: 7184
TA: 10
WU: 44

6. AUTHOR(S)
William P. Janson*, Judi E. See*, Joseph T. Riegler*, Iris Davis*, Gilbert G. Kuperman

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
* Logicon Technical Services, Inc.
PO Box 317258
Dayton, OH 45437-7258

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory, Human Effectiveness Directorate
Crew System Interface Division
Air Force Materiel Command
Wright-Patterson AFB, OH 45433-7022

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-HE-WP-TR-1999-0003

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
The current investigation examined automatic target cueing (ATC) and target localization performance with first generation forward-looking infrared (FLIR) imagery collected during the Theater Missile Defense Eagle Smart Sensor ATC (TESSA) flight demonstration program at Eglin Air Force Base, FL. Sixteen observers viewed 360 dynamic FLIR images that varied in terms of aiding (unaided versus aided), ATC accuracy (50% hits versus 75% hits), the amount and type of background clutter ("open," "treeline," and "sparse" sites), and slant range to the target array (8 km, 6 km, and 4 km). All images contained the target to be detected, a transporter-erector-launcher (TEL), as well as two support vehicles. After viewing each FLIR image sequence, participants identified the location of the TEL and rated their confidence in their decision. The results revealed that ATC cueing enhanced operators' confidence in their decision making but did not alter their localization accuracy or perceptual sensitivity (d') relative to the unaided condition, an outcome which may be attributable to ceiling effects (mean correct localizations were 99.4%). In addition, operators performed somewhat more slowly with the cueing than without, though the difference was small. Further, ATC accuracy did not affect any aspect of operator performance or confidence, again due to ceiling effects. The high level of performance effectiveness in this study indicates a need to examine the effects of ATC assistance at ranges greater than 8 km.

14. SUBJECT TERMS
Aided & Unaided Target Acquisition, Automatic Target Recognition, ATR, Automatic Target Cueing, ATC, FLIR, Signal Detection Theory, Theater Missile Defense, TMD, Operator Performance, Human Performance, Mobile Missiles, Time Critical Targets, TESSA

15. NUMBER OF PAGES
54

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UNL

This page intentionally left blank.

PREFACE

This effort was conducted by the Information Analysis and Exploitation Branch, Crew System Interface Division, Human Effectiveness Directorate of the Air Force Research Laboratory (AFRL/HECA), Wright-Patterson Air Force Base, OH, under Work Unit 71841044, "Crew-Centered Aiding for Advanced Reconnaissance, Surveillance, and Target Acquisition." It was supported by Logicon Technical Services, Inc. (LTSI), Dayton, Ohio, under Contract F41624-94-D-6000, Delivery Order 0007. Mr. Don Monk was the Contract Monitor.

The authors gratefully acknowledge the personnel of the Air Force Theater Missile Defense Attack Operations Program Office of the Aeronautical Systems Center (ASC/FBXT) and the Air Force Research Laboratory's Sensors Directorate (AFRL/SN) for providing the imagery, automatic target cueing (ATC) and ground truthing declaration data, and supporting information.

The authors also wish to acknowledge the vigorous efforts put forth by several LTSI coworkers: Mr. Robert Stewart both for ongoing project management support and technical advice; Mr. Luther Storey for his invaluable assistance in reviewing and preparing the imagery files; Mr. David Robinow for the development of the experimental software controlling image presentation and data collection; and Ms. Elisabeth Fitzhugh for her superb job editing and formatting the report.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table **Page**

SECTION 1. INTRODUCTION

Background

Countering the threat posed by theater missiles (TMs) has become a high U.S. defense priority since Operation Desert Storm. Throughout this conflict, significant defense resources were expended protecting our allies from missile strikes with intercept and destroy systems (e.g., Patriot Advanced Capabilities - 2 [PAC-2] being among the most effective). Locating and targeting mobile missile launchers such as the Scud-B transporter-erector-launcher (TEL) prior to or post-launch proved to be a difficult task for air defenses. In fact, Fulghum (1994) reports a lack of evidence for the destruction of a single Scud TEL during the Gulf War. These experiences led to the expansion of the U.S. Theater Air Defense (TAD) to include not only aircraft defense, but also defense against TMs and their supporting infrastructure.

During typical battlefield operations, conventional short-range missiles are not high priority targets, and in small numbers can be insignificant (Mumma & Bell, 1993). However, when these missiles are armed with chemical, biological, or nuclear warheads, attack operations to destroy them prior to launch are of the highest priority. The current concept of operations (CONOPS) for command and control (C^2) against time critical targets (TCTs) calls for a rapid response, given the high priority of TCTs and the relatively short 10 - 15 minute window of attack opportunity (Jones, 1997). Consequently, a core Air Force objective is to attack and destroy TMs and other TCTs as far into the enemy's territory as possible, preferably, prior to launch, when they are the least threatening.

In order to meet TCT mission objectives, a superior level of connectivity and integration is required between air and spaceborne sensors, C^2 nodes, and attack platforms to provide a comprehensive, "fused" representation of the battlespace to all command levels. Emerging technologies that advance the level of integration between these defense operations are currently being emphasized by the U.S. Department of Defense (DoD). One such technology that is of interest to the present report is the application of automatic target recognition (ATR) and cueing (ATC) technologies to sensor imagery to improve operator capability to locate, track, identify, and engage mobile ground targets.

In broad terms, ATR/ATC has been defined as the "computer processing of image data from optical, radar, infrared, or other imaging sensors to identify image locations that correspond to specific physical objects (targets)" (Augustyn, 1992; p. 105). However, in more specific terms, ATC has been defined as the "automated detection (and possible classification) of an object of possible military interest, while ATR refers to the automated recognition (and possible identification) of a detected object" (Kuperman, 1997, p. 38). The application of ATR/ATC technologies to the military target search and identification problem has received considerable attention in recent years. In the military domain, an important application of ATR/ATC technology is to improve the performance and survivability of attack aircraft by providing a means for quick, accurate, and automated detection of targets in radar images (Delashmit, 1989). In this scenario, real-time automated processing of imagery could provide the decision-making operator with information ranging from areas believed to contain a target of interest to the actual identification of a specific target. Both ATC and ATR technologies are also suggested to offer the potential to greatly reduce the operator workload involved in military attack operations.

While automaticity is the ultimate goal of ATR/ATC technologies, the U.S. House Permanent Select Committee on Intelligence has advocated that current efforts in the ATR/ATC domain be directed at *assisting*, rather than replacing, the operator (Aerospace Daily, 1996). The rationale for a near-term emphasis on "assisted target recognition" systems is primarily technology-driven. In the near-term, operators are envisioned as remaining "in-the-loop" to facilitate the cognitive aspects of the target recognition process and to make final targeting decisions, rendering the human a critical component of any automated target acquisition system.

The TESSA Program

In support of the development and testing of ATR/ATC technologies, the Theater Missile Defense (TMD) Attack Operations (AO) System Program Office (ASC/FBXT) sponsored a data collection effort during March and April 1995 at Eglin Air Force Base, Florida. This data collection effort has become known as the *T*heater Missile Defense *E*agle *S*mart *S*ensor and *A*TC (TESSA) program. The primary objective of the TESSA program was to collect both medium resolution synthetic aperture radar (SAR) imagery and high quality digital first generation forward-looking infrared (FLIR) imagery of mobile missile targets for use in TMD targeting simulations, laboratory investigations, and in the development of ATR/ATC algorithms. The

SAR and FLIR imagery were collected with the F-15E APG-70 radar and the Low Altitude Navigation and Targeting Infrared for Night (LANTIRN) targeting pod sensor systems, respectively. For a more thorough description of each sensor system refer to See, Riegler, Fitzhugh, & Kuperman (1996). FLIR systems such as LANTIRN complement the APG-70 radar by providing the aircraft with a day/night, under the weather, low altitude, air to ground capability (Goble, Williams, Pratt, Wald, Rubin, & Hanson, 1980). While the APG-70/LANTIRN combination itself should enhance target acquisition performance beyond that achieved with either system alone, the addition of ATC/ATR technology should serve to further improve the operator's ability to detect, track, identify, and attack mobile missile threats.

Mission Description

Both the SAR and FLIR imagery were collected during a total of nine missions flown at various times of the day and night across three different clutter sites. The flight profiles for the data collection missions were identical. Each mission consisted of ten passes toward an array of three stationary vehicles described below. The flight profile for the first pass was initiated at 74 km (40 nautical miles [nmi]) from the vehicles, and at 37 km (20 nmi) on each subsequent pass. On each pass, the angle of approach to the targets, which were always aligned towards magnetic north, also varied systematically. On the first pass, the approach angle was 135°; on the remaining passes, approach angle varied from 180° (south, tail-on view) to 0° (north, head-on view) in intervals of 22.5°, resembling a half wagon wheel. A nominal air speed of 420 knots true ground speed (KTGS) was maintained, resulting in data collection passes of approximately 3 minutes from 37 km to overflight.

Each data collection pass was divided into a SAR portion and a FLIR portion. The SAR data collection portion, flown at 17,000 feet, began at pass initiation and continued to 18.5 km (10 nmi) from the target site. At this point, the aircraft descended to 10,000 feet and initiated the LANTIRN FLIR portion of data collection, which continued to overflight. At the beginning of the FLIR portion, the LANTIRN targeting pod was cued to the pre-briefed target location. The display was monitored when the target array became detectable, at approximately 12 km (6.5 nmi), whereupon the task was to slew the targeting pod to the innermost vehicle in the array. The sensor display itself consisted of a narrow field of view of the approaching scene. The display symbology included a centrally-located white crosshair and flight-related parameters (such as slant range and Z-time [Greenwich Mean Time]) which appeared as white alphanumeric

symbols along the periphery of the display. Each pass was recorded both on 8 mm analog cassette and on digital tape. The 8 mm tape imagery captured the entire pass (both SAR and FLIR) commencing with SAR mapping, while the digital imagery contained the FLIR portion only.

Vehicles

The target array consisted of a Scud-B mobile missile TEL (13.0 m long by 3.2 m wide by 3.4 m high), a ZiL-131 communications van (6.9 m long by 2.4 m wide by 2.4 m high), and a German MAN 4-axle all-wheel drive truck (8.9 m long by 2.5 m wide by 3.0 m high) carrying a high pressure air compressor (HIPAC) unit (see Figure 1). The TEL, which served as the primary target of interest, was an authentic and fully functional (except for its inert and unfueled missile) specimen of a late 1960's Soviet battlefield mobile tactical missile launcher. The ZiL was a 3-axle all-wheel drive unit widely used throughout the former Soviet Union in a variety of military roles. In the TESSA program, the ZiL served as a command and control vehicle that could be expected to accompany the TEL to an unprepared launch site. The MAN was included in the target array as a "confuser" target, for it possessed many of the same features as the TEL, including size, number of axles, type of drive, and engine location. While all three vehicles were present during all the TESSA missions, the arrangement of the vehicles varied from mission to mission.

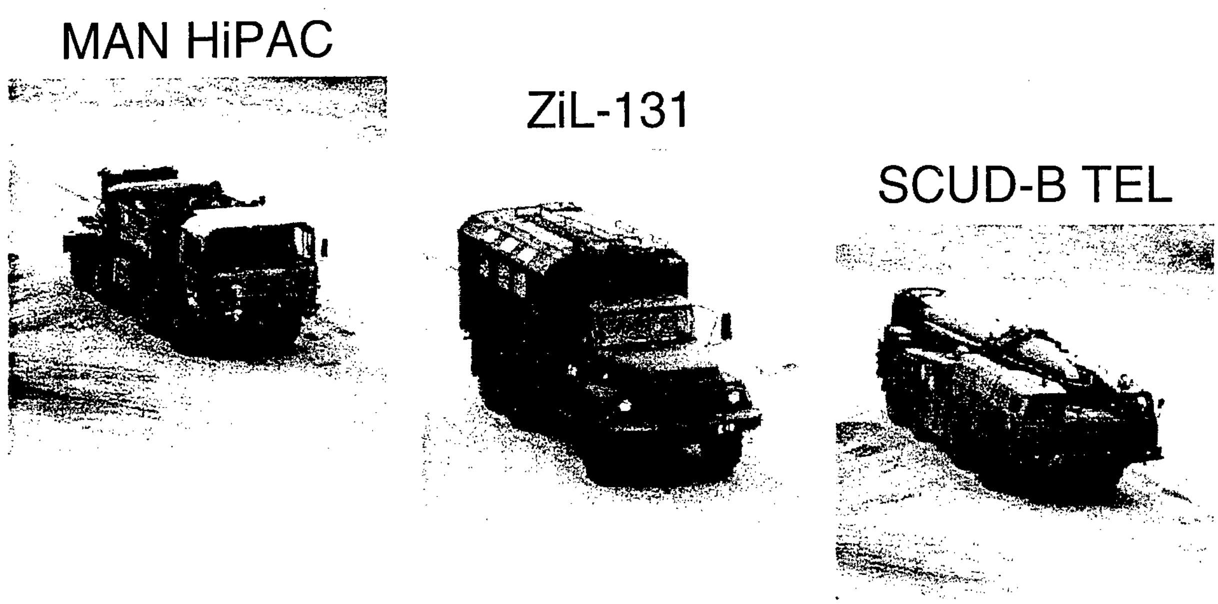


Figure 1. Vehicles in the TESSA missions.

Clutter Sites

The TESSA missions were flown over three distinct geographical locations at Eglin Air Force Base. These locations varied with respect to the amount and type of ground cover that might act as clutter sources to the SAR or FLIR sensor systems and were referred to as the open, treeline, and sparse sites. The open site, which contained primarily sandy soil and a combination of low cut vegetation and grass, was selected to represent the lowest level of clutter-object confusion to the FLIR sensor. The treeline site represented a moderate level of clutter. The treeline formed a homogeneous background that was relatively inconspicuous in comparison to the open fields and roads in the foreground. In both the open and treeline sites, the targets were always located in the open, on or near the roads. The sparse site consisted of a flat grassy plain containing widely spaced trees and bushes. Among the three sites, the sparse site provided the highest level of clutter to the FLIR sensor; however, no action was taken to mask or conceal targets amid the clutter.

FLIR Target Detection Performance

Prior to the design and development of this experiment, personnel in the Crew Aiding and Information Warfare Analysis Laboratory (CIWAL) reviewed the 8 mm tape imagery of the various TESSA missions to determine video quality and to identify primary variables that could potentially affect target detection performance. As a result of this process, some TESSA missions were discarded due to poor image quality and sporadic provision of ATC information. In the remaining missions, the potential variables selected for inclusion in a study of operator target acquisition performance were background clutter, range bin (distance from the target array), and ATC availability.

Background Clutter

The background clutter in which a target is located can affect FLIR target detection and recognition performance. The infrared signature of a target generally can be more easily detected if the object is situated in a low clutter region with minimal vegetation, as opposed to a more highly cluttered scene characterized by both greater amounts and more varied types of ground cover (Strzempko & Pritchard, 1990). High background clutter also provides more confuser objects that can be mistaken for targets, necessitating scrupulous examination of each object before arriving at a target decision (Shumaker, 1979). The resultant reduction in target

salience relative to the background can prolong visual search time and increase the likelihood that the target signature will be missed. Increasing the amount and type of vegetation can also degrade performance effectiveness. Highly cluttered scenes may contain background objects with signatures similar to that of the target, increasing the potential for false alarms—that is, the incorrect designation of nontarget objects as targets (Rotman, Kowalczyk, Cartier, & Chang, 1994).

The detrimental effects of high background clutter on FLIR target acquisition performance have been shown empirically. In a previous study using unaided FLIR imagery from the TESSA program, See et al. (1996) had observers view several seconds of dynamic imagery before determining whether a crosshair symbol on the display was positioned over the TEL target. Results indicated that the observers' ability to discriminate the TEL from the ZiL and MAN support vehicles was higher for the open site than for the sparse site (d' index of perceptual sensitivity values of 3.6 and 2.9, respectively). In a similar study, Beideman, Gomer, and Levine (1980) examined observers' ability to detect and recognize three types of military ground vehicles in three levels of background clutter (low, medium, and high). Subjects were required to locate the target vehicle in the scene, as well as to determine its type. Results indicated that at an initial slant range of 30,000 feet (9 km), response times for detection and recognition were significantly longer in medium and high clutter scenes as compared to those with no ground clutter. In addition, observers were able to detect and recognize targets at significantly greater distances overall in the low clutter scenes, as opposed to the performance degradation shown in both the medium and high clutter scenes (medium and high clutter performance results were similar). Taken together, these studies demonstrate the potentially degrading effects of background clutter on operator target detection and recognition performance. It remains to be seen whether the addition of ATC information moderates the detrimental effects of clutter on FLIR target acquisition performance.

Range Bin

A second factor which can affect target acquisition performance is the slant range from the target array. This distance is typically represented in one kilometer intervals referred to as range bins (e.g., a range bin of 8 would correspond to a distance of 9 km to 8 km from the target). Range becomes an important factor because sensor imagery and displays must provide sufficient image quality in order to permit target identification beyond the effective ranges of anti-aircraft

defenses (Beideman et al., 1980). Consequently, it is important to establish the range at which effective target acquisition performance can be achieved with current sensors. Several previous investigations have found that target detection and recognition performance are sensitive to variations in target range (Beideman et al., 1980; See et al., 1996; Turner, 1995; Valeton & Bijl, 1994). In general, these studies suggest that performance decrements in detection, as well as recognition, can be expected as the distance from the target array increases. However, the precise range at which performance accuracy will reach an acceptable level can vary depending on the type of target to be detected and the background clutter in which it occurs.

ATC Availability

A third factor which may affect FLIR target acquisition performance is the availability of ATC information. Currently, only a small percentage of the articles published in the ATR/ATC domain have addressed human-machine interface (HMI) issues or have evaluated operator performance with ATR/ATC systems, (see Toms & Kuperman, 1991, for a review). Interest has primarily centered on ATR/ATC presentation format (Adams, 1991), algorithm reliability (Weisgerber & Savage, 1990), and operator opinions regarding the utility of autoclassifiers for targeting (Kibbe, Adams, Weisgerber, & Savage, 1990). In general, these studies suggest that operators favor the integration of ATR and sensor information and that target acquisition performance is generally improved relative to unaided conditions. While these systems are relatively good at detecting targets, they tend to have high false alarm rates, which can reduce operator confidence in the information. Recent data suggests that, in order for ATC information to be effective, accuracies of at least 70% must be achieved, with false alarms minimized to four or fewer per image (Becker, Hayes, & Gorman, 1991; Fulkerson, 1980; Jauer, Quinn, Hockenberger, & Eggleston, 1986; Kibbe & Weisgerber, 1991; Weisgerber & Savage, 1990). In fact, a recent operator performance study utilizing SAR imagery suggests that if all ATC cues are false alarms (a distinct possibility), operator performance is actually worse than if no aiding information is provided at all (See, Davis, & Kuperman, 1997). Similarly, Carr (1988), maintains that while ATC information can be beneficial in enhancing human search performance, this advantage may be limited to cueing only real targets (i.e., keeping cues to a minimum). Consequently, the mere presentation of ATC information is not sufficient, in and of itself, to improve operator target acquisition performance; rather, the ATC information must also be highly reliable or operator confidence may be lost, resulting in operator indifference to the information or even a reduction in operator performance.

Currently, there is a need for empirical studies to provide data which could assist ATC developers in defining useful methods of integrating ATC information with sensor imagery and in identifying conditions under which ATC-generated information would be of most use to aircrews. At present, such information appears to be lacking in the current literature. The present study was conducted, in part, to fill this void.

The Theory of Signal Detection

In order to examine the effects of cueing, ATC accuracy, background clutter, and range on operator performance in the present study, the techniques of the Theory of Signal Detection (TSD) were applied. TSD is a model of perceptual processing that is frequently used to characterize performance effectiveness in target acquisition tasks (Gescheider, 1985; Green & Swets, 1966; Macmillan & Creelman, 1991; See & Kuperman, 1995; See et al., 1996; See, Warm, Dember, & Howe, 1997; Wilson, 1992). The application of TSD to a target detection task entails the derivation of two independent measures of performance: perceptual sensitivity (d') and response bias (c). The d' index of sensitivity is a perceptual measure that provides a bias-free estimate of the observer's ability to discriminate targets from nontargets. The index of response bias, c, provides an independent assessment of the operator's general willingness to make a detection ("target") response, which can vary on a continuum from conservative to lenient. Both measures are derived from observers' hits (correct detections) and false alarms (errors of commission) during the course of a task. A TSD analysis is preferable to separate examinations of hits and false alarms because it permits performance to be characterized *independently* in terms of sensing abilities and decision making processes with measures that simultaneously take both the hits and false alarms into account, as reflected in the computing formulae for sensitivity and bias:

$$\mathbf{d'} = \mathbf{z}_{\mathbf{H}} - \mathbf{z}_{\mathbf{FA}} \quad [1]$$

$$\mathbf{c} = -.5\left(\mathbf{z}_{\mathbf{H}} + \mathbf{z}_{\mathbf{FA}}\right) \quad [2]$$

In each formula, z represents the standard normal deviate associated with proportions of hits (H) and false alarms (FA), both of which enter directly into the derivation of each TSD index.

In many detection tasks in which TSD is applied, including target acquisition, observers may be required not only to detect the presence of a target (e.g., noting a trio of objects that are grouped in a pattern typical of a TEL and its supporting vehicles), but also to determine its precise location (i.e., target designation). Thus, once they have determined that a target is present, observers must decide which of the several "target-like" objects in the scene has the greatest likelihood of being the target. The probability of correctly determining the target's location when it is present is derived from the joint probability of making both a correct detection of the target and a correct identification of its location. Similarly, in still other tasks the objective may not be to determine whether or not a target is present, but rather to decide where it is located, given that the target is always present. Under these types of circumstances, it is still possible to apply TSD and obtain estimates of operator sensitivity, with some modification. The d' index of sensitivity for target localization can be interpreted as the operator's ability to differentiate the actual target from other alternative "target-like" objects that may be present. It is estimated, from either a computational formula or tables of d', on the basis of the number of alternatives available for designation and the operator's ensuing proportion of correct localization responses (Hacker & Ratcliff, 1979; Macmillan & Creelman, 1991). Since response bias in target localization tasks tends to be neutral, its calculation is often not necessary. However, if desired, the index of bias can be obtained to provide a measure of the observer's degree of caution or conservatism in making the localization response. The calculation is the same as that for target detection, with the proportion of correct localizations substituted for hits (Macmillan & Creelman, 1991).

Purpose

The purpose of this experiment was to evaluate the effect of ATC information on human target acquisition performance with first generation FLIR imagery. Utilizing FLIR imagery and ATC information from the TESSA flight scenarios, observers were tasked to view several seconds of FLIR imagery before locating the TEL target. The variables of interest consisted of background clutter, target range, cue condition, and cue accuracy. Background clutter consisted of the open, treeline, and sparse sites from the TESSA flights. Target range, or the distance between the aircraft sensor and the target, was represented by 3 one kilometer range bins: 5 to 4 km, 7 to 6 km, and 9 to 8 km. Cue condition involved the presence (aided) or absence (unaided) of ATC cue boxes overlaid on the imagery to assist in TEL identification. Cue accuracy

represented the precision of placement of the cue boxes in those scenes in which they appeared. In this study, the two levels of accuracy investigated were 50% precision in cue box placement (ATC designation of the TEL) and 75% precision. The imagery selected for the experiment represented a subsample of the total imagery collected during the TESSA missions. Input selection for the study was based on image quality of the scene, accuracy of ATC information provided, and representation across the variables of interest in this experiment.

SECTION 2. METHOD

Experimental Design

The basic design consisted of a 2 (cueing) x 2 (cue accuracy) x 3 (clutter) x 3 (range) mixed design. The within-subjects independent variables consisted of cue condition (aided, unaided), background clutter (open, treeline, and sparse sites), and target range (4, 6, and 8 km). Nested within the aided cue condition was a between-subjects variable consisting of cue accuracy (50%, 75%). Participants were randomly assigned to each level of cue accuracy. The imagery for each cueing condition was blocked, and its presentation order balanced across participants. Within each block, background clutter and target range were randomly presented. The dependent variables consisted of the percentage of correct TEL localizations, *d'*, response time, and confidence rating.

Apparatus and Stimuli

The study was conducted in the Crew Aiding and Information Warfare Analysis Laboratory (CIWAL) located within the Air Force Research Laboratory at Wright-Patterson Air Force Base, Ohio. The imagery was presented on a Silicon Graphics O^2 color graphics system (Model # W10-195S-4G64V), including a 19 in. monitor and a mouse for target localization. A keypad was also used to initiate each trial and to enter confidence ratings. Fifty-six unique "pass" scenarios representing variations in background clutter, target range, approach angle, and vehicle configuration were presented across conditions (validated by available ground truth; Pryce, 1995). Approach angle and vehicle configuration were not included as task variables due to limited data. Each "pass," as presented to the operator in this study, consisted of FLIR imagery for a one kilometer range bin depicting a three second approach towards the TEL, MAN, and ZiL. No nontarget scenes were presented. Each pass began at the far edge of the range bin and proceeded towards the target area. A pass was comprised of 90 individual files or frames of digitized imagery. Each frame measured 8.25"L x 3.875"H (720 x 356 pixels) on the display and subtended a visual angle of approximately 10°. To portray a dynamic presentation during the trial, these frames were presented in sequence at a nominal rate of 30 frames per second.

For the aided condition trials, two boxes representing ATC information were overlaid on the imagery. The ATC algorithm was developed by Hughes Aircraft Company and represented a model-based vision algorithm with low fidelity thermal models. The models were based on Computer Aided Design (CAD) geometric models painted to simulate different surface temperatures. Internal characteristics were not modeled in either the geometry or the thermal models. The ATC algorithm was applied in the laboratory to the recorded TESSA FLIR imagery at the conclusion of the flight. The monitor brightness and contrast settings were preset and held constant throughout the study, facilitated by taking weekly luminance readings of the 50% and 100% white segments of a 16 point gray scale. On average, the mean and standard deviation luminance values for these segments across the 5 weeks of data collection was 14.668 cd/m^2 (*SD* = 0.425) for 50% white and 48.881 cd/m^2 (*SD* = 2.145) for 100% white.

Participants

Sixteen males served as test participants. They ranged in age from 25 to 50 years (*M* = 38.1 years, *SD* = 6.8 years). While particular emphasis was placed on recruiting weapon systems operators, their limited availability led to subsequent recruitment of pilots, navigators, and other active duty military personnel. The final background composition of these observers included four weapon systems operators, four navigators, four pilots, and four non-rated active duty military personnel. Half of these participants had direct operating experience with FLIR sensors (ranging from 3 to 1000 hours), while two others had viewed FLIR imagery previously in CIWAL experiments. Fifteen of these individuals possessed a corrective visual-acuity of 20-20; and one, 20-25.

Procedure

Upon arrival to the test facility, the participant was led to a crew briefing room and given the consent form. Following consent, the individual completed a brief background questionnaire and received a description of the study from the experimenter. Visual acuity was then tested using the Snellen chart. Participants were permitted to wear corrective eyewear during the testing. Following the acuity test, the participant was taken to an isolated work area for data collection. The observer was seated before the O^2 monitor and given the task instructions, followed by six practice trials (three for each cue condition). During practice, the experimenter

described the scenario and answered questions. In the event the participant needed more time to gain familiarity with the imagery, the practice trials were repeated. Once practice was completed, the observer was ready for data collection. The experimenter left the room and turned-off the overhead lights. The only remaining room light emanated from a small lamp directed towards the corner of the room, illuminating the experimental area while avoiding glare on the monitor.

To begin a trial, the observer depressed a "Ready" key on the keypad. This initiated three seconds of dynamic FLIR imagery presentation showing a continuous approach to a target area containing the TEL, MAN, and ZiL. Then, and without interruption, the final frame of this imagery remained on the display for a maximum of eight seconds while the observer located the TEL among the three vehicles. Localization of the target was accepted only during the static portion of this presentation, during which time a "*Designate Target*" message appeared on the display beside the imagery. Localization was accomplished by using a mouse to position a cursor over the center of the TEL (or any object believed to be the TEL) and clicking the upper-left mouse button. In the aided trials, the ATC boxes were overlaid on the image during the static presentation. Once the eight seconds elapsed, or the observer localized a target, the scene vanished from the display and was replaced by an "*Enter Confidence Rating*" message. The observer responded by depressing one of six numeric keys on the keypad to reflect his level of confidence in his TEL identification. Low confidence was reflected by entering a "1" or "2;" medium confidence, a "3" or "4;" and high confidence, a "5" or "6." Given the limited number of unique images and the necessity of repeating each image several times throughout the study, observers were asked to base their confidence ratings on the quality of the imagery, rather than on familiarization effects due to repetition of images. In the event the observer failed to record a localization during the static image presentation time frame, a confidence rating of "1" was entered for the trial. Once a confidence value was entered, a "*Ready*" command appeared on the screen signifying readiness for the next trial. Figure 2 provides a graphical representation of a trial sequence.

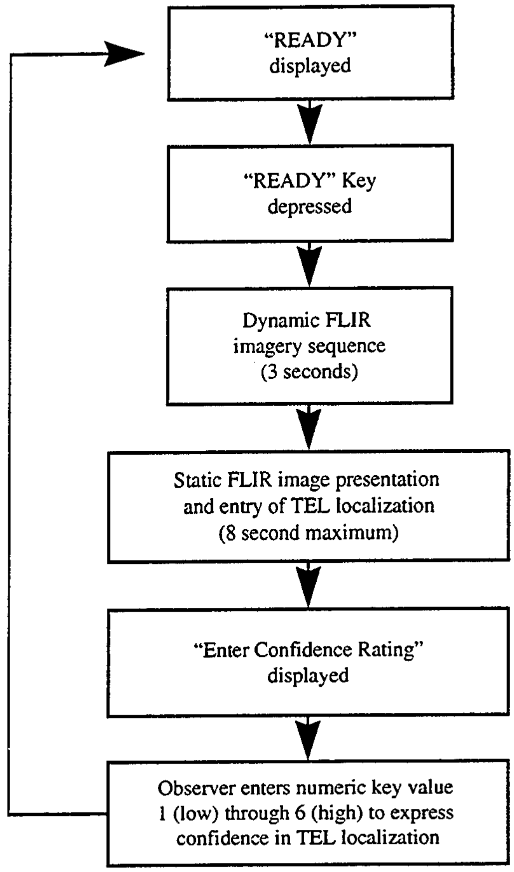


Figure 2. Sequence of events for a trial.

As mentioned previously, the static image portion of the aided trials included the presentation of two boxes overlaid on the image. Two boxes were presented to model the multiple ATC reports which typically occurred during each TESSA pass. The precise image locations for the box overlays were derived from actual TESSA ATC reports. While the original ATC information was available sporadically throughout each TESSA mission and pass, this experiment used only the ATC reports taken from the final frame in each range bin. This was done in order to ensure consistency in both initiation and duration of ATC output presentation within and across range bins, and also explains the reason for overlaying the boxes during the static image portion of the trial (representative of the final frame of the range bin). The accuracy of box placement on the TEL was either 50% or 75%, depending upon observer group. Observers were informed that they could either accept or reject the cued information; their primary goal was to identify and designate the center of the TEL. For the unaided trials, the image presentation was identical except for the absence of the ATC boxes.

Each participant completed 360 experimental trials (180 aided and 180 unaided), with rest breaks permitted at any time between trials. Figure 3 depicts the distribution of these trials for the two cue accuracy groups. Once a participant had finished all 360 trials, he completed a questionnaire regarding the aided FLIR imagery. The experimenter also documented any additional comments made regarding the study. This completed the data collection session. The average session length was approximately 68 min. (SD = 9 min.).

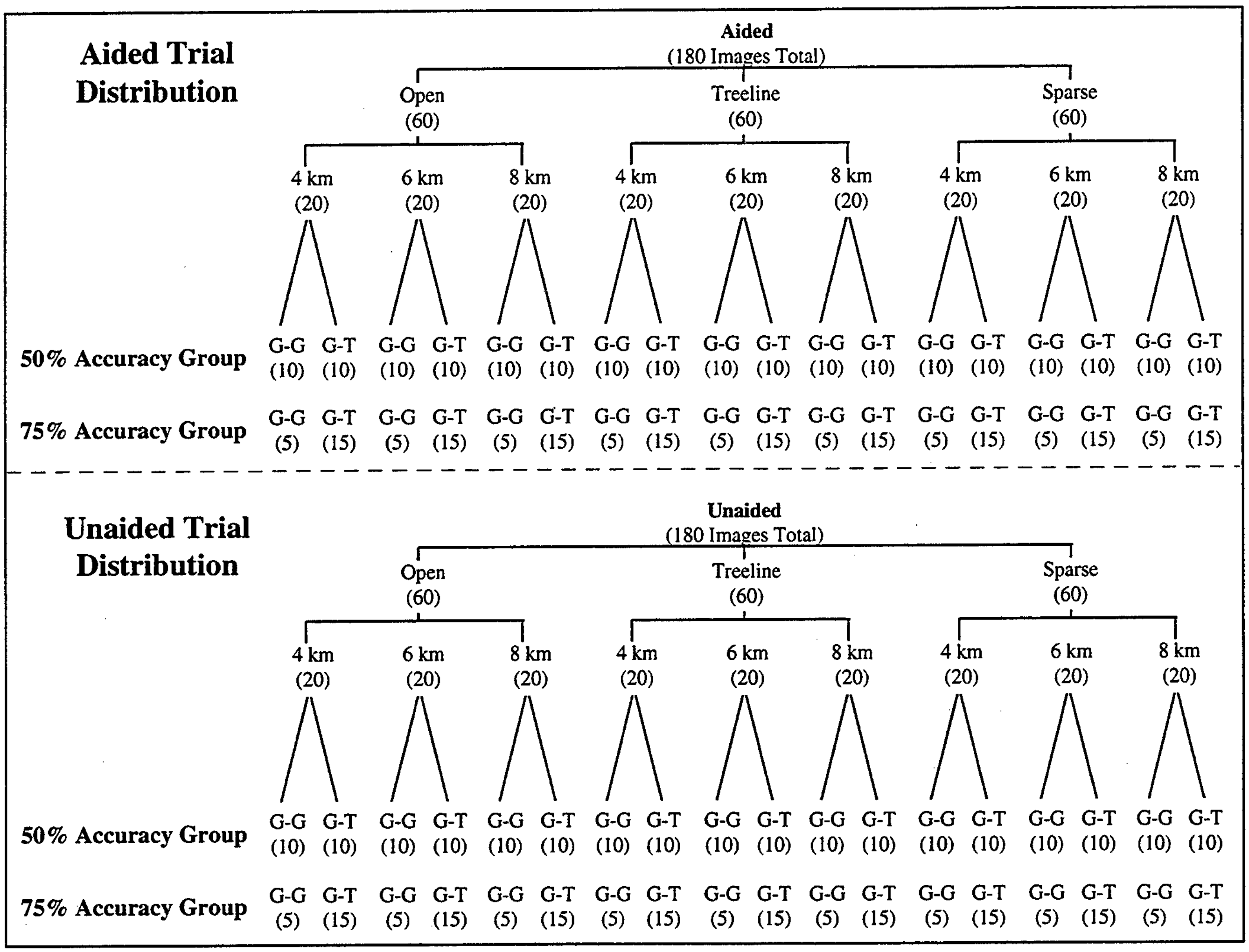


Figure 3. Trial distribution for each level of cueing and accuracy group. G–G denotes both ATC boxes within the trial appeared on the ground, while G–T denotes one ATC box appeared on the ground, the other on the TEL.

SECTION 3. RESULTS

We examined performance effectiveness via four primary dependent variables: the percentage of correct localizations, perceptual sensitivity (d'), reaction time (RT) for correct localizations, and operators' confidence ratings. Although operators were instructed to designate the center of the TEL, the correctness of localization included an error tolerance based on the length of the TEL target in the unique image used. Two factors— variations in TEL size across trials (a function of range) and the time restriction imposed by the task—supported allowance for some degree of leniency in acceptable TEL localizations. If the operator's localization point lay no further than one half TEL-length from the center of the target, it was considered correct. Specifically, the maximum acceptable distance value for classifying a correct TEL localization was based on the TEL's width and height, using the following formula:

$$\text{Acceptable Distance} = \sqrt{\left(\frac{width}{2}\right)^2 + \left(\frac{height}{2}\right)^2} \qquad [3]$$

This formula was used because it inherently accommodated some degree of imprecision in manually designating the TEL. It also ensured that a designation located anywhere on the TEL (front, center, rear) would be considered correct.

Percentages of correct localizations were then used to derive the d' index of perceptual sensitivity for each individual in the various experimental conditions by consulting the appropriate tables of d' for localization (Hacker & Ratcliff, 1979; Macmillan & Creelman, 1991). For localization tasks such as ours, the primary determinant of perceptual sensitivity (in conjunction with the percentage of correct localizations) is the number of alternative items that can be selected as the target. As in previous studies using imagery from the TESSA program (Davis, See, Shacklett, & Kuperman, 1996; See, Davis, & Kuperman, 1997), we operated under the assumption that the three vehicles in the target array represented the three alternatives that were available for possible selection as the TEL target. Before determining d', percentages of 0 and 100 were first mathematically adjusted by means of a procedure recommended by Snodgrass and Corwin (1988) where a value of 0.5 was added to each frequency and divided by N + 1 (where N represents the total number of trials).

Two preliminary analyses of the data were conducted. First, *t*-tests indicated that imagery block presentation order (aided first versus unaided first) had no effect on any of the performance measures, $p > .05$. Second, inspection of each participant's data during the course of the experiment seemed to suggest that the manipulation of ATC accuracy (75% versus 50%) did not affect performance. The results of 2 (accuracy) x 3 (clutter) x 3 (range) analyses of variance (ANOVAs) for each dependent variable in the aided condition confirmed that there were no main or interactive effects associated with ATC accuracy, $p > .05$. Hence, we decided to exclude ATC accuracy from further analysis. The results that follow were derived from 2 (cueing) x 3 (clutter) x 3 (range) repeated measures ANOVAs, disregarding ATC accuracy. The alpha level for all ANOVAs was set at .05. Probabilities for any effect containing three or more levels (i.e., clutter and range) were obtained via the Huynh-Feldt epsilon adjustment (Huynh & Feldt, 1970, 1976).

Percentage of Correct Localizations

Mean percentages of correct localizations at each clutter site and range in the aided and unaided conditions are presented in Table 1. First, the overall mean of 99.4% indicates that the operators' performance in this study was almost perfect. Second, the figures in the table reveal essentially no difference between the aided and unaided conditions. With respect to clutter, performance accuracy did decline somewhat from the open site as compared to performance in the treeline and sparse sites. Finally, within each site, performance was most accurate at a range of 4 km and least accurate at a range of 8 km.

The ANOVA of the means in Table 1 revealed significant main effects for clutter, $F(2, 30) = 3.86$, $p < .03$, and for range, $F(2, 30) = 4.44$, $p < .05$. However, the effect for cueing was not significant, $p > .05$. Of the possible interactions, only the Clutter x Range interaction was significant, $F(4, 60) = 3.40$, $p < .02$. For the main effects of clutter and range, post hoc correlated *t*-tests were used to determine where the significant differences lay. The overall alpha for each set of tests was .20, yielding an alpha of .07 for each individual comparison. The results of the post hoc analyses appear in Tables 2 and 3. The presence of an asterisk implies that the two conditions under comparison were significantly different at the individual alpha of .07, whereas "NS" signifies that any differences were not significant. As can be seen in Table 2, performance accuracy was significantly better in the open site than in the treeline and sparse

sites, while performance in treeline and sparse sites was approximately equal. With respect to Table 3, the percentage of correct localizations was higher at 4 km ($M = 99.9$, $SD = 0.5$) than at 6 km ($M = 99.6$, $SD = 1.3$) and at 8 km ($M = 99.7$, $SD = 3.7$) but results for the 6 km did not differ statistically from those for the 8 km conditions.

Table 1. Mean Percentage of Correct Localizations (Standard Deviations in Italics) at Each Clutter Site and Range for the Aided and Unaided Conditions.

Clutter Site	Range	Condition: Aided	Condition: Unaided	Mean
Open	4	100.0	100.0	100.0
		0.0	*0.0*	*0.0*
	6	100.0	99.7	99.8
		0.0	*1.2*	*0.9*
	8	99.7	99.7	99.7
		1.2	*1.2*	*1.2*
	Mean	99.9	99.8	99.8
	SD	*0.7*	*1.0*	*0.9*
Treeline	4	100.0	100.0	100.0
		0.0	*0.0*	*0.0*
	6	99.4	99.1	99.2
		1.7	*2.0*	*1.8*
	8	99.4	97.5	98.4
		1.7	*6.3*	*4.6*
	Mean	99.6	98.8	99.2
	SD	*1.4*	*3.9*	*2.9*
Sparse	4	100.0	99.7	99.8
		0.0	*1.2*	*0.9*
	6	99.7	100.0	99.8
		1.2	*0.0*	*0.9*
	8	97.5	98.4	98.0
		4.1	*4.4*	*4.2*
	Mean	99.1	99.4	99.2
	SD	*2.7*	*2.6*	*2.6*
Overall Mean		99.5	99.3	99.4
SD		*1.8*	*2.8*	*2.3*

Table 2. Results of Post Hoc Correlated t-Tests of Correct Localizations for Clutter Site.

	Treeline	Sparse
Open	*	*
Treeline		NS

Table 3. Results of Post Hoc Correlated t-Tests of Correct Localizations for Range.

	6 km	8 km
4 km	*	*
6 km		NS

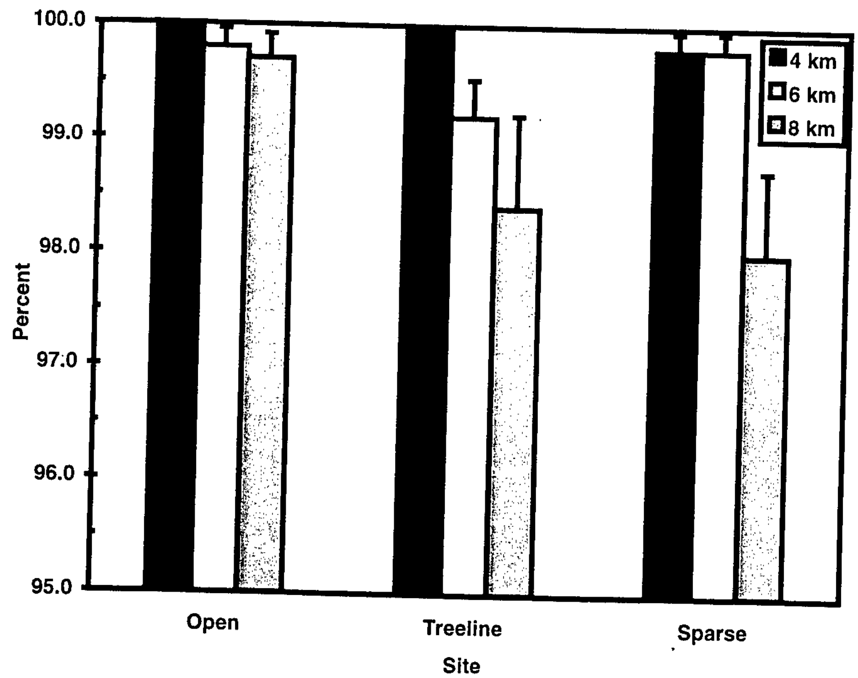


Figure 4. Mean percentage of correct localizations at each clutter site and range (error bars represent the standard error of the mean).

The two-way interaction between clutter and range is depicted in Figure 4. As can be seen in the figure, the percentage of correct localizations tended to remain stable across variations in range bin at the open site but not at the treeline and sparse sites, where deteriorations in performance accuracy were evident. In order to assess the statistical significance of the effects of range within each clutter site, post hoc correlated *t*-tests were conducted. Specifically, differences in the percentages of correct localizations among the three ranges were compared separately within each level of clutter. The overall alpha for the set of

tests was .20, producing an alpha of .02 for each of the nine individual comparisons. The results of these analyses revealed no differences in the percentage of correct localizations among the three ranges in the low clutter (open) site. In the medium clutter (treeline) site, performance was significantly more accurate at 4 km as compared to 6 km. In the high clutter (sparse) site, the percentage of correct localizations was significantly greater at 6 km as compared to 8 km. No other differences were statistically significant.

Perceptual Sensitivity

Mean perceptual sensitivities (d') were calculated for each experimental condition by using the observed mean percentages of correct localizations and consulting the appropriate tables of d' provided by Hacker & Ratcliff (1979). The mean d' scores at each clutter site and range for the aided and unaided conditions appear in Table 4 on the following page. Recall that these values represent a bias-free estimate of the observer's ability to discriminate the TEL from other alternative "target-like" objects in the scene. According to guidelines provided by Craig (1984), d' scores of about 3.5 are indicative of a very easy task. Thus, as with the overall percentage of correct localizations, the mean d' score in the study (~ 3.2) indicates that the target localization process was relatively easy. With respect to the condition of cueing, the means in Table 4 show that sensitivity was similar in the aided and unaided conditions. The figures in Table 4 also reveal that sensitivity was greater in the open site than in the treeline and sparse sites. Within each site, d' tended to decline as the range from the target increased from 4 km to 8 km.

The ANOVA of the d' scores revealed significant main effects for clutter, $F(2, 30) = 5.20$, $p < .01$, and for range, $F(2, 30) = 4.96$, $p < .04$. The effect for cueing was not significant, $p > .05$. The only interaction to attain statistical significance was the Clutter x Range interaction, $F(4, 60) = 4.35$, $p < .005$. For the main effects of clutter and range, post hoc correlated *t*-tests were used to determine where the significant differences lay. The overall alpha for each set of tests was .20, yielding an alpha of .07 for each individual comparison. The post hoc tests for clutter indicated that sensitivity was greater in the open site than in the treeline and sparse sites, which did not differ from each other. With respect to range, perceptual sensitivity was greater at a range of 4 km ($M = 3.24$, $SD = 0.02$) than at 6 km ($M = 3.21$, $SD = 0.06$) and 8 km ($M = 3.13$, $SD = 0.18$), where it did not differ.

Table 4. Mean Perceptual Sensitivity (Standard Deviations in Italics) at Each Clutter Site and Range for the Aided and Unaided Conditions.

Clutter Site	Range	Condition: Aided	Condition: Unaided	Mean
	4	3.25	3.25	3.25
		0.00	*0.00*	*0.00*
Open	6	3.25	3.22	3.23
		0.00	*0.14*	*0.07*
	8	3.25	3.22	3.23
		0.00	*0.14*	*0.07*
	Mean	3.25	3.23	3.24
	SD	*0.00*	*0.06*	*0.03*
	4	3.25	3.25	3.25
		0.00	*0.00*	*0.00*
Treeline	6	3.18	3.15	3.16
		0.18	*0.22*	*0.13*
	8	3.18	3.04	3.11
		0.18	*0.48*	*0.24*
	Mean	3.20	3.14	3.17
	SD	*0.08*	*0.21*	*0.10*
	4	3.25	3.22	3.23
		0.00	*0.14*	*0.07*
Sparse	6	3.22	3.25	3.23
		0.14	*0.0*	*0.07*
	8	3.00	3.10	3.05
		0.38	*0.41*	*0.29*
	Mean	3.15	3.19	3.17
	SD	*0.13*	*0.14*	*0.10*
Overall Mean		3.20	3.19	3.20
SD		*0.05*	*0.13*	*0.06*

The nature of the Clutter x Range interaction is portrayed graphically in Figure 5. As can be seen in the figure, the decline in sensitivity as range increased was more pronounced in the treeline and sparse sites than in the open site, where sensitivity remained more or less stable as the range from the target varied. As in the case of correct localizations, the statistical significance of the effects of range within each clutter site was assessed by means of post hoc correlated *t*-tests, with an overall alpha of .20 for the set of comparisons (alpha of .02 for each individual comparison). As expected, the results of these analyses revealed no differences in sensitivity among the three ranges at the open site. At the treeline site, sensitivity was greater at 4 km than at 6 km. At the sparse site, sensitivity was greater at 6 km than at 8 km. No other differences were statistically significant.

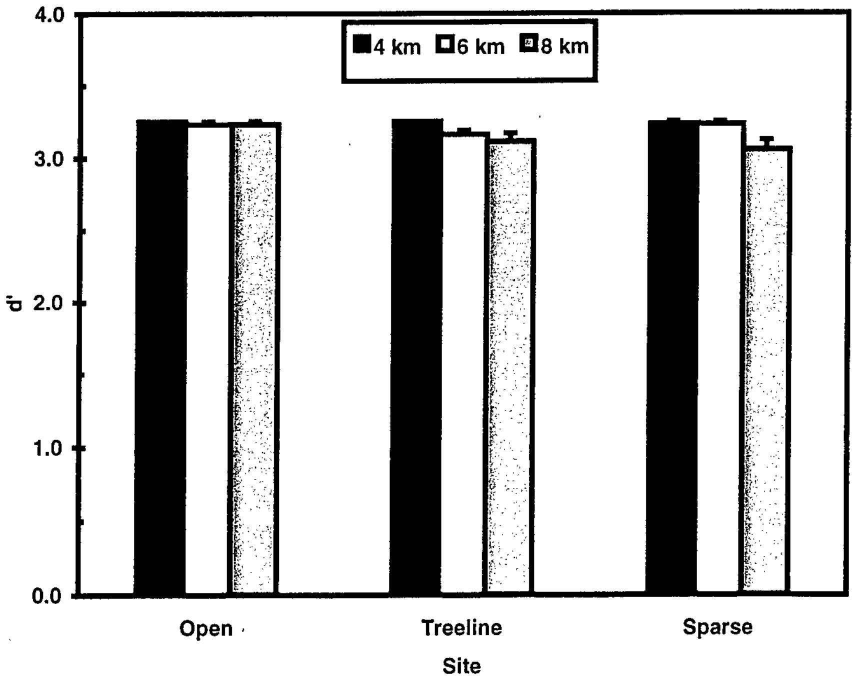


Figure 5. Mean perceptual sensitivity at each clutter site and range (error bars represent the standard error of the mean).

The d' values from Table 4 were also used to derive "notional ROC curves." These curves were based on the notion that localizing the TEL from the three target array was likened to "detecting" the TEL, and that every such selection was either correct or incorrect. Failure to correctly select the TEL was indicative that the MAN, ZiL, or some non-target object was selected. As the experiment did not specifically distinguish among the non-target responses, all incorrect selections were treated the same. Consequently, all correct localizations were treated as "detections," while all incorrect localizations were treated as "false alarms." It was further assumed that, for any given experimental condition, the mean percentage of correct localizations (still expressed as a fraction) represented an estimate of the probability of detection, P_d. The P_{fa}, on the other hand, based on the premise that any selection that was not a detection was a false alarm, was represented as:

$$\mathbf{P_{fa} = 1 - P_d}$$

This equation represents an essential feature for the derivation of the notional ROC curves for it provides the points (P_{fa}, P_d) necessary for calculating an "equivalent, yes - no, d'." Given that $z(P)$ denotes the standard normal deviate corresponding to P, then the desired equivalent d' can be written as:

$$\mathbf{d'_{eq} = z(P_d) - z(P_{fa})}$$

From this d'_{eq} it is possible to construct a corresponding ROC by using the simple TSD model with various threshold levels. But there is a serious caution. The resulting ROC will not give any other points that are possible outcomes of the experiment. In a conventional ROC, P_d and P_{fa} not only approach 0.0 together, they also approach 1.0 together. Any possible outcome of the present experiment, however, satisfies only the relation that $P_d + P_{fa} = 1$. Hence, when P_d is one, P_{fa} is zero, and vice versa. Therefore, only the point at which d'_{eq} is calculated legitimately characterizes operator performance as indicated by the experiment.

Table 5 presents the corrected mean percentage of correct localizations from the tabulation of perceptual sensitivities. These figures were obtained by working backward through the tables by Hacker and Ratcliff, entering the tabulated values of d' (Table 4) in the column for $M = 3$, and reading off the corresponding probability.

Table 5. Adjusted Probabilities of Correct Localization (from Hacker & Ratcliff, 1979).

Background	Range (km)	Aided	Unaided
Open	4	0.980	0.980
Open	6	0.980	0.979
Open	8	0.980	0.979
Treeline	4	0.980	0.980
Treeline	6	0.977	0.976
Treeline	8	0.977	0.971
Sparse	4	0.980	0.979
Sparse	6	0.979	0.980
Sparse	8	0.969	0.974

From these adjusted probabilities, the new, equivalent index (d'_{eq}) was computed (Table 6), and represented by the "notional ROC curves" depicted in Figures 6 through 11. Of note, is the observation that these d'_{eq} values fall near a value of 4, representing almost perfect performance (Craig, 1984 guidelines).

Table 6. Equivalent Perceptual Sensitivity, d'_{eq} values (Simple TSD model).

Background	Range (km)	Aided	Unaided
Open	4	4.107	4.107
Open	6	4.107	4.067
Open	8	4.107	4.067
Treeline	4	4.107	4.107
Treeline	6	3.991	3.955
Treeline	8	3.991	3.791
Sparse	4	4.107	4.067
Sparse	6	4.067	4.107
Sparse	8	3.733	3.886

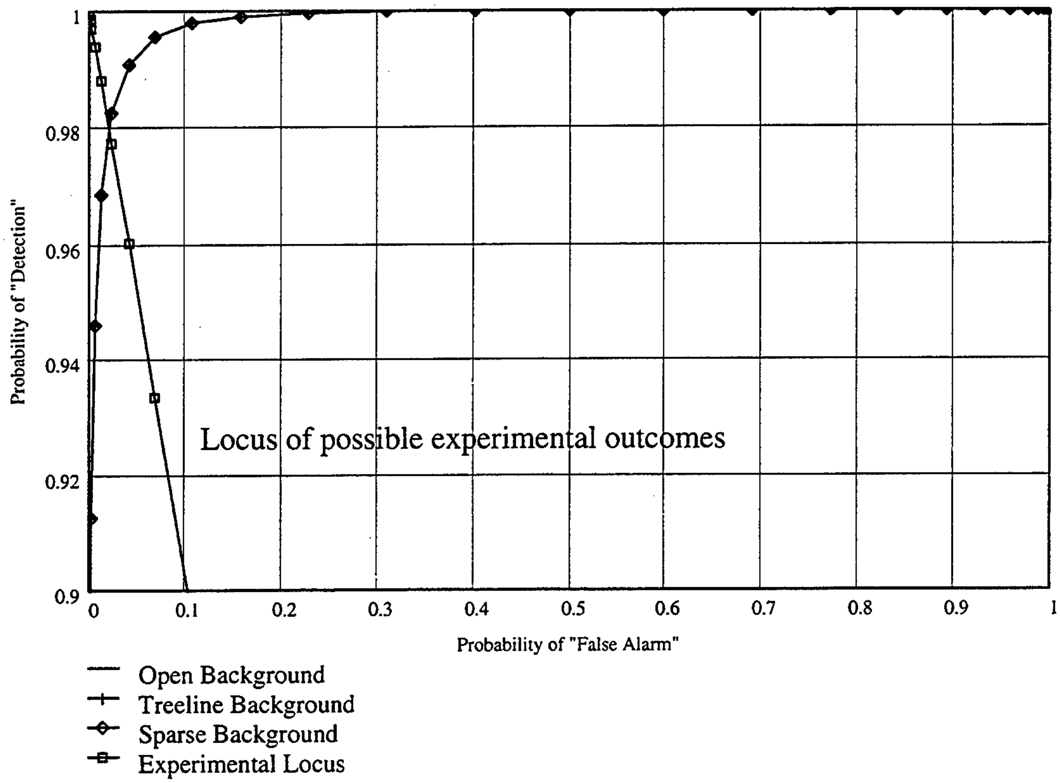


Figure 6. Notional ROC at 4 km, Aided—Note: Expanded P_d scale.

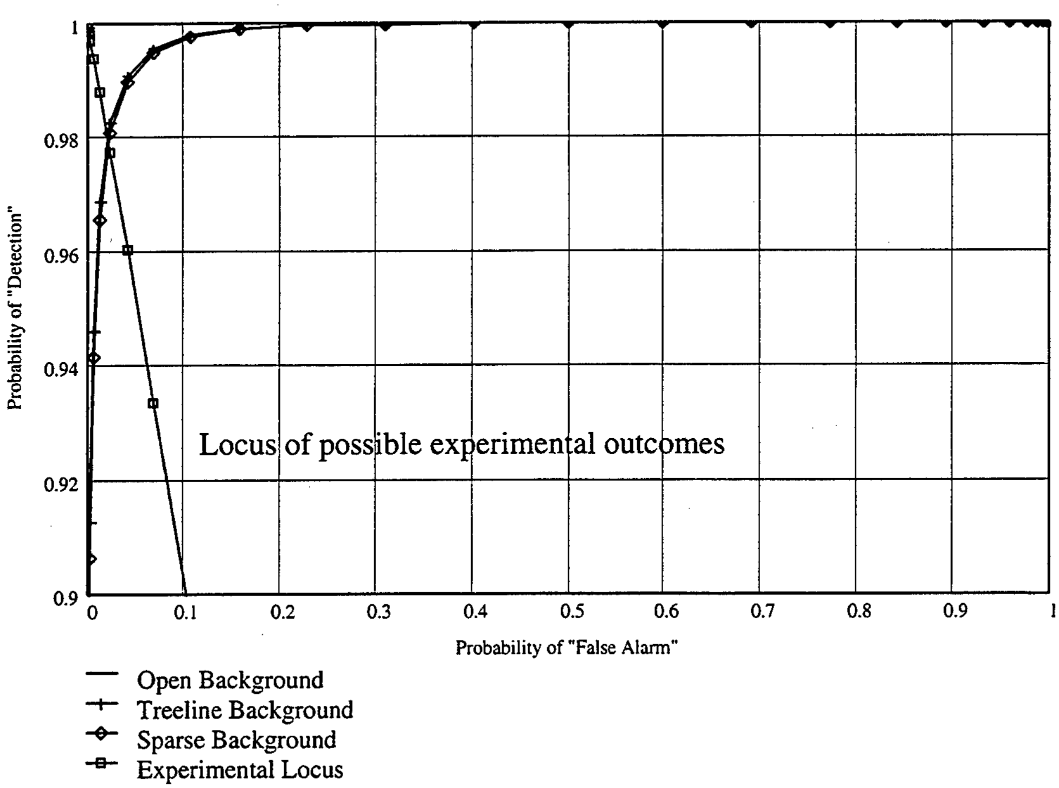


Figure 7. Notional ROC at 4 km, Unaided—Note: Expanded P_d scale.

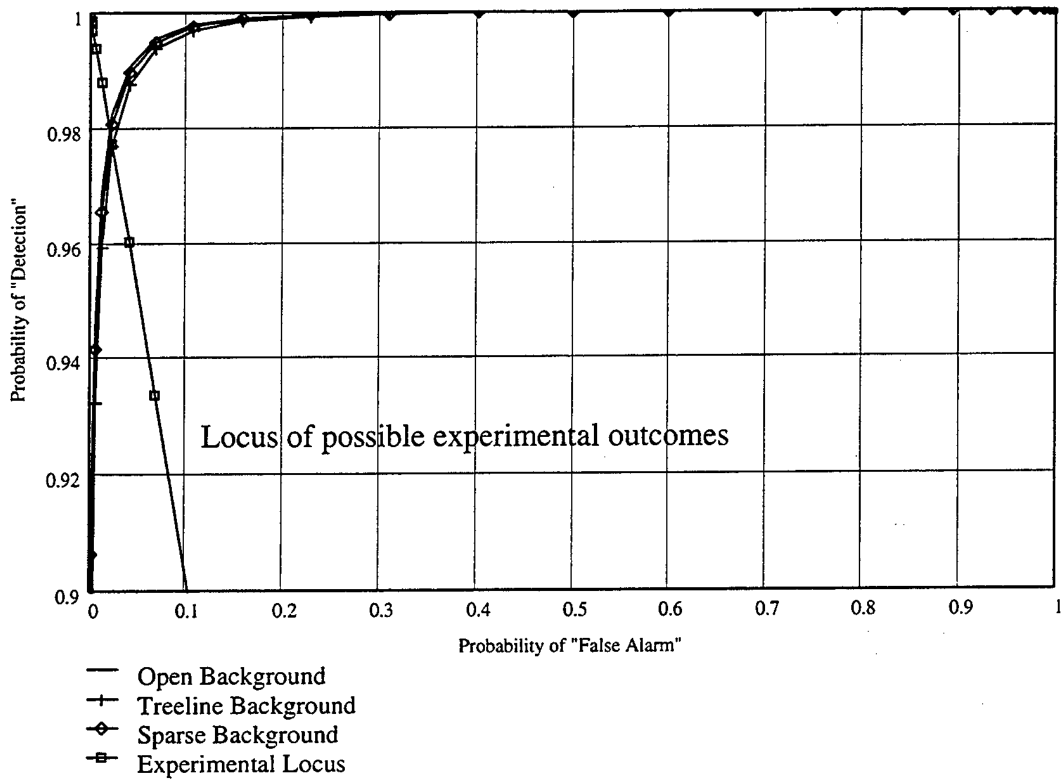


Figure 8. Notional ROC at 6 km, Aided—Note: Expanded P_d scale.

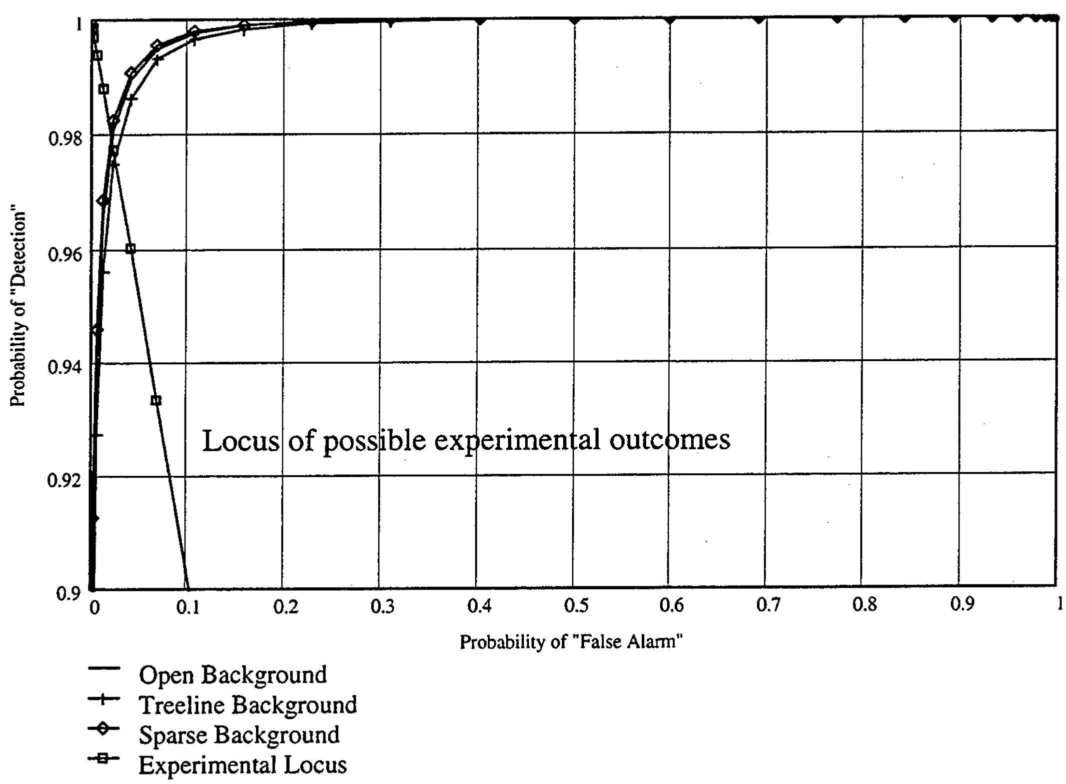


Figure 9. Notional ROC at 6 km, Unaided—Note: Expanded P_d scale.

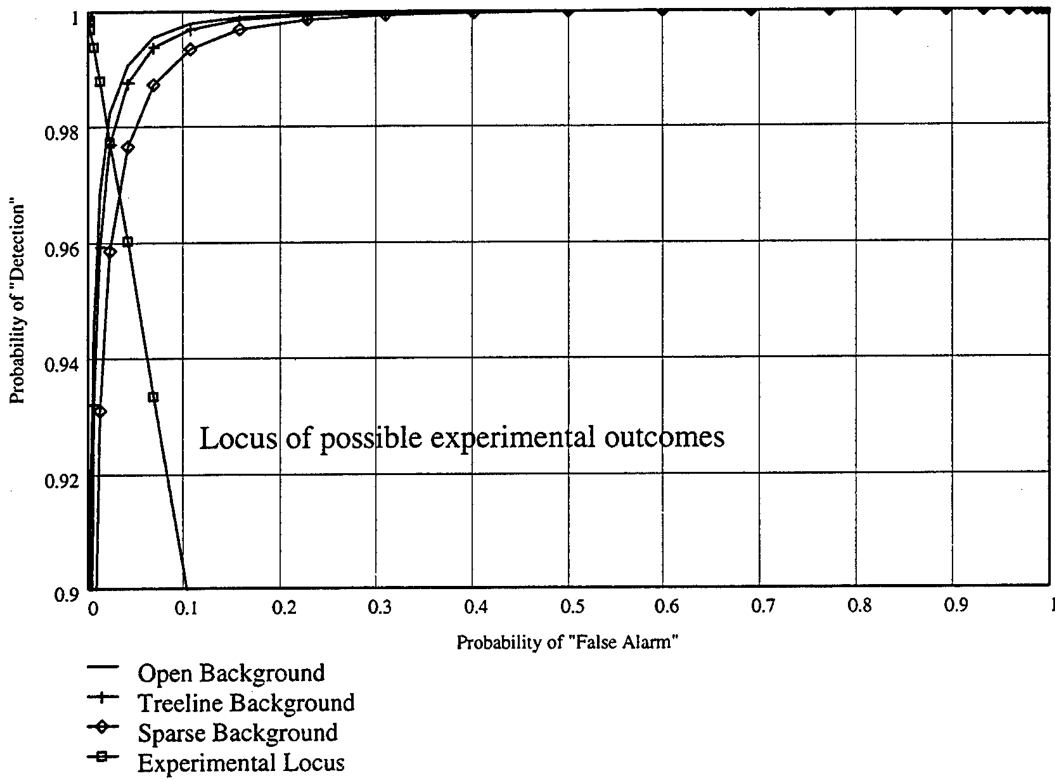


Figure 10. Notional ROC at 8 km, Aided—Note: Expanded P_d scale.

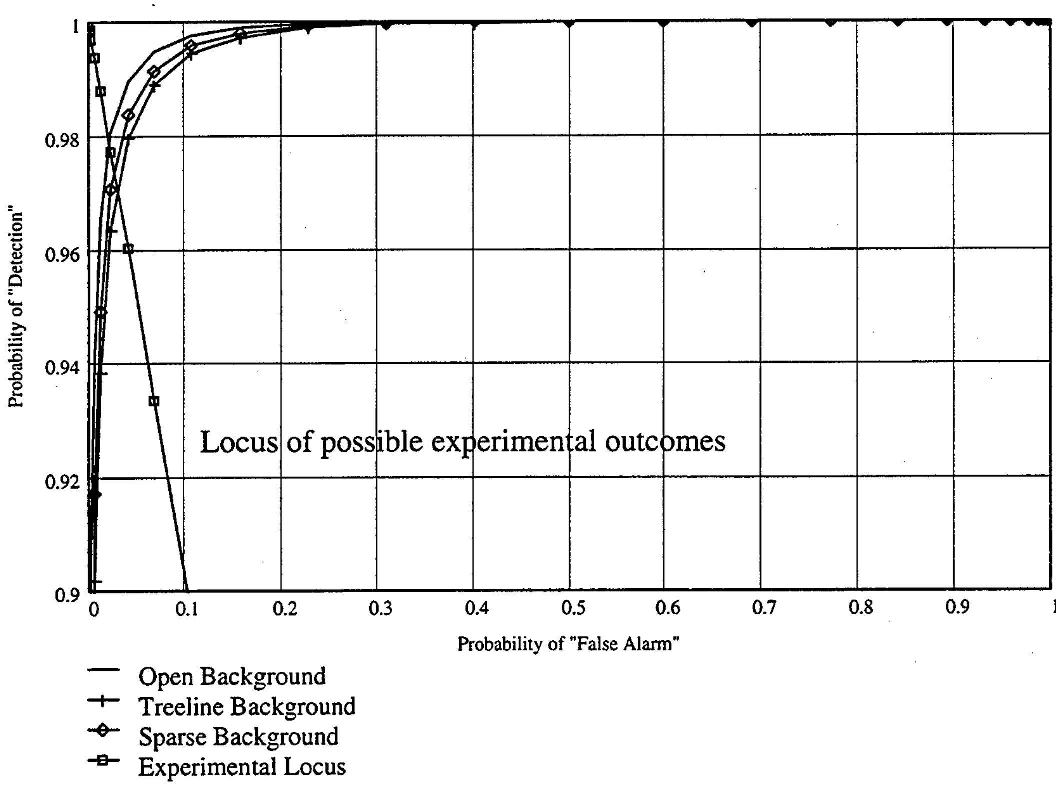


Figure 11. Notional ROC at 8 km, Unaided—Note: Expanded P_d scale.

These graphs of "notional ROCs" are a different, and somewhat unconventional, way of presenting the information contained in the tables of this report. Their use, however, may be justified by the fact that although performance was excellent for all cases (approximately 97% correct and 3% incorrect localizations for all experimental conditions) the graphs clearly show the beginnings of deterioration at the 8 km range, regardless of cueing level. They also show substantial indications of performance differences across the three clutter sites. It is expected that these trends would have continued had it been possible to extend the experimental study to ranges greater than 8 km. The intersections of the "ROCs" with the locus of possible experimental operating points provide exactly the same information as the adjusted fractions of correct localizations listed in Table 5. That is, the intersections provide the same information as do the tables of the report, and are, in fact, the significant points of the "ROCs." The intersections and the tabular data, therefore, support exactly the same conclusions.

Reaction Time for Correct Localizations

Reaction time (RT) for correct localizations was defined as the time (in seconds) from onset of the static image presentation (i.e., the beginning of the eight second interval) until the observer localized the TEL with the mouse button. The mean RT scores (in seconds) at each clutter site and range for the aided and unaided conditions appear in Table 7. With respect to the condition of aiding, the means in Table 7 indicate that observers were somewhat slower in the aided condition than in the unaided. With respect to clutter, operators responded most quickly in the open site as compared to the treeline and sparse sites. Further, within each site, the RT for correct localizations tended to increase as the range from the target increased.

The ANOVA of the RT scores revealed significant main effects for cueing, $F\ (1, 15) = 17.37, p < .0008$; clutter, $F\ (2, 30) = 15.94, p < .0001$; and range, $F\ (2, 30) = 51.12, p < .0001$. None of the interactions was statistically significant, $p > .05$. For the main effects of clutter and range, post hoc correlated *t*-tests were performed using an overall alpha of .20 for each set of tests (an alpha of .07 for each individual comparison). The *t*-tests for clutter indicated that operators were indeed faster in the open site than in the treeline and sparse sites, where the RT for correct localizations did not differ. Post hoc tests further revealed that RT deteriorated progressively as the range increased from 4 km ($M = 1.1$, $SD = 0.3$) to 6 km ($M = 1.3$, $SD = 0.5$) to 8 km ($M = 1.7$, $SD = 0.6$).

Table 7. Mean RT (in seconds) for Correct Localizations (Standard Deviations in Italics) at Each Clutter Site and Range for the Aided and Unaided Conditions.

Clutter Site	Range	Condition: Aided	Condition: Unaided	Mean
Open	4	1.1	0.9	1.0
		0.3	*0.3*	*0.3*
	6	1.4	1.1	1.2
		0.4	*0.4*	*0.4*
	8	1.7	1.5	1.6
		0.7	*0.6*	*0.6*
	Mean	1.4	1.2	1.3
	SD	*0.6*	*0.5*	*0.5*
Treeline	4	1.3	1.1	1.2
		0.3	*0.3*	*0.3*
	6	1.5	1.3	1.4
		0.5	*0.4*	*0.5*
	8	2.0	1.6	1.8
		0.6	*0.5*	*0.6*
	Mean	1.6	1.3	1.5
	SD	*0.6*	*0.4*	*0.5*
Sparse	4	1.3	1.1	1.2
		0.3	*0.3*	*0.3*
	6	1.4	1.2	1.3
		0.4	*0.4*	*0.5*
	8	1.9	1.7	1.8
		0.6	*0.5*	*0.6*
	Mean	1.5	1.3	1.4
	SD	*0.5*	*0.5*	*0.5*
Overall Mean		1.5	1.3	1.4
SD		*0.5*	*0.5*	*0.5*

Confidence Ratings

Operator confidence ratings were derived from the numerical values, ranging from 1 through 6, used to describe the observer's level of confidence in correctly localizing the TEL during each trial. A value of "1" represented low confidence, while a value of "6" represented high confidence. Mean confidence ratings at each clutter site and range for the aided and unaided conditions are presented in Table 8. As can be seen in the table, overall confidence was fairly high, averaging 4.8 on a scale that ranged from 1 to 6. Confidence ratings were higher in the aided condition than in the unaided condition. In addition, confidence ratings were higher in the open site than in the remaining two sites. Finally, within each site, confidence tended to decline as the range from the target increased, an effect that was more prominent in the treeline and sparse sites.

The ANOVA of the confidence ratings indicated significant main effects for cueing, $F\ (1, 15) = 8.58, p < .01$; clutter, $F\ (2, 30) = 20.89, p < .0001$; and range, $F\ (2, 30) = 105.86, p < .0001$. Of the interactions, the Cueing x Range interaction was significant, $F\ (2, 30) = 8.24, p < .006$, as was the Clutter x Range interaction, $F\ (4, 60) = 15.79, p < .0001$. Post hoc correlated *t*-tests were conducted for the main effects of clutter and range, with an overall alpha of .20 for each set of comparisons and an alpha of .07 for each individual comparison. As expected, based on the means in Table 8, confidence ratings were significantly higher in the open site than in the treeline and sparse sites, where confidence did not differ. Tests for range showed that confidence diminished progressively as range increased from 4 km ($M = 5.7$, $SD = 0.4$) to 6 km ($M = 5.0$, $SD = 0.8$) to 8 km ($M = 3.9$, $SD = 1.0$).

Table 8. Mean Confidence Rating (Standard Deviations in Italics) at Each Clutter Site and Range for the Aided and Unaided Conditions.

Clutter		Condition		
Site	Range	Aided	Unaided	Mean
	4	5.9	5.9	5.9
		0.2	*0.2*	*0.2*
Open	6	5.3	5.0	5.2
		0.6	*0.7*	*0.7*
	8	4.5	4.1	4.3
		0.9	*0.8*	*0.9*
	Mean	5.2	5.0	5.1
	SD	*0.8*	*1.0*	*0.9*
	4	5.8	5.7	5.7
		0.4	*0.4*	*0.4*
Treeline	6	4.9	4.7	4.8
		1.0	*0.9*	*0.9*
	8	3.8	3.2	3.5
		1.2	*1.0*	*1.1*
	Mean	4.8	4.5	4.7
	SD	*1.2*	*1.3*	*1.3*
	4	5.6	5.4	5.5
		0.5	*0.5*	*0.5*
Sparse	6	5.2	4.8	5.0
		0.8	*0.8*	*0.8*
	8	4.1	3.6	3.8
		1.0	*0.9*	*1.0*
	Mean	5.0	4.6	4.8
	SD	*1.0*	*1.1*	*1.0*
Overall Mean		5.0	4.7	4.8
SD		*1.0*	*1.1*	*1.1*

The interaction between cueing and range is portrayed in Figure 12. As shown in the figure, differences in confidence between the aided and unaided conditions were negligible at a range of 4 km, becoming more pronounced only at the more distant ranges of 6 km and 8 km. Post hoc correlated *t*-tests were used to test for significant differences between the aided and unaided conditions at each range. The overall alpha for the set of tests was .20, producing an alpha of .07 for each individual comparison. The results of those tests confirmed that confidence in the aided and unaided conditions did not differ at 4 km; however, at 6 km and 8 km, confidence was significantly greater in the aided condition.

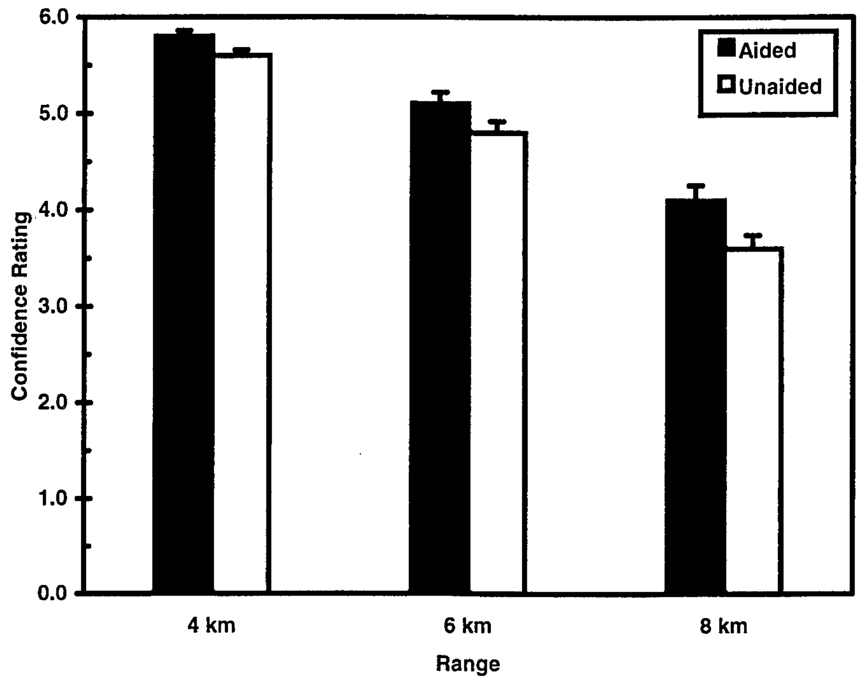


Figure 12. Mean confidence rating in the aided and unaided conditions at each range bin (error bars represent the standard error of the mean).

Finally, the Clutter x Range interaction is portrayed graphically in Figure 13. At 4 km, confidence tended to decline as the amount of background clutter increased from the open to the sparse site. At 6 km and 8 km, however, confidence appeared to be the lowest in the treeline site. Post hoc correlated *t*-tests were used to determine whether there were differences in confidence among the three sites within each range bin. The overall alpha was set at .20, yielding an alpha of .02 for each individual comparison. The results of those tests are depicted in the alphabetic labels in Figure 13. Within each range bin, dissimilar labels indicate significant differences between the respective sites. Thus, at 4 km,

confidence declined significantly only in the sparse site (as compared to the open site). At 6 km, confidence declined from the open site to the treeline and sparse sites, which did not differ from each other. Finally, at 8 km, confidence was highest in the open site and lowest at the treeline site.

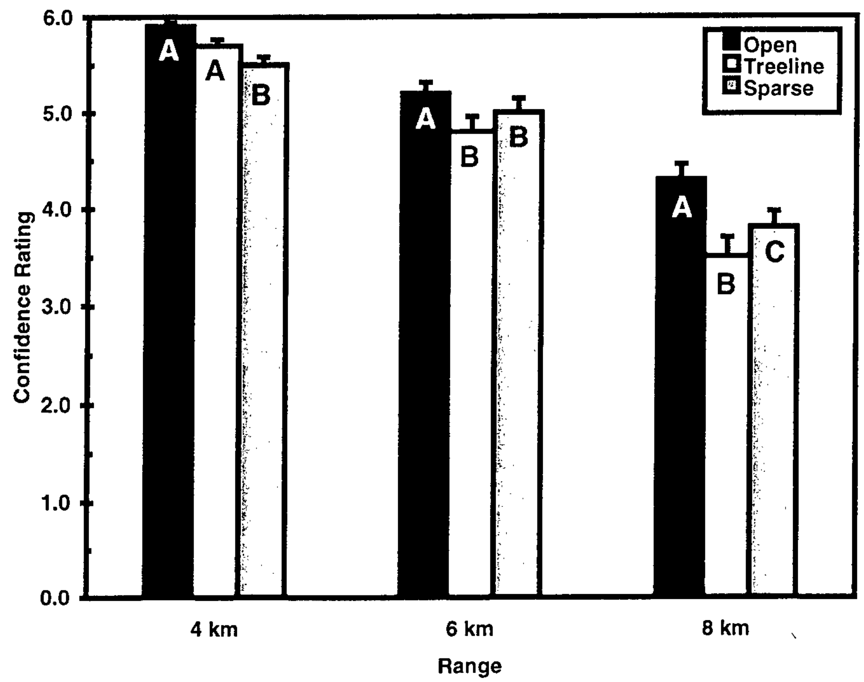


Figure 13. Mean confidence ratings in the open, treeline, and sparse clutter sites at each range bin (error bars represent the standard error of the mean).

Summary of Performance Results

In summary, statistical analyses of the data revealed that there were no differences between the high and low ATC accuracy conditions during the aided portion of the study. Consequently, our main analyses consisted of 2 (cueing) x 3 (clutter) x 3 (range) repeated measures ANOVAs of the percentage of correct localizations, *d*', RT for correct localizations, and confidence ratings. Those analyses revealed that the effects of cueing could be seen only in the RT and confidence rating data—operators were more confident in the aided condition than in the unaided condition—but they were also somewhat slower in their target-localization responses. Neither the percentage of correct localizations nor perceptual sensitivity was affected by the cueing. Finally, the effects of clutter and range were statistically significant in all analyses. Operators performed more accurately, faster, and with greater confidence in the open site than in the treeline and sparse sites, where no differences were observed in any of the dependent variables. With respect to range, both the percentages of correct localizations and *d*' were

significantly higher at a range of 4 km as compared to 6 km and 8 km. Further, RT increased and confidence decreased progressively as the range increased from 4 km to 8 km.

Questionnaire Results

Following the data collection session, we asked each participant to complete a questionnaire to assess their reactions to the ATC information. The first question inquired whether the *ATC* was helpful in target localization. The responses were equally divided; eight participants thought the ATC cue box was helpful and the other eight thought it was not. Further, the responses within the low and high accuracy groups were also equally divided among positive and negative groups. Participants who found the ATC to be helpful indicated that it increased their confidence in their decision or that it assisted them in locating the target array, obviating the need to search the entire scene. (Note: see the Appendix for a detailed listing of questionnaire responses). Participants who did not think the ATC was useful commented that they had already decided where the target was when the cue boxes appeared and, therefore, did not find the information helpful. They also commented that the apparent low accuracy of the system hindered their decision-making.

The second question asked participants to estimate the *accuracy* of the ATC. Responses from the low (50%) accuracy group ranged from 20% to 70% (M = 48%, SD = 15), whereas those from the high (75%) accuracy group ranged from 10% to 70% (M = 51%, SD = 21). The mean estimates of the low and high accuracy groups were not statistically different, $t\,(14) = .28$, $p > .05$.

The third question asked operators to provide an estimate of what they would consider an *acceptable level of accuracy* for an ATC. Responses from the low accuracy group ranged from 70% to 95% (M = 84%, SD = 7). Responses from the high accuracy group ranged from 75% to 95% (M = 89%, SD = 7). As with the estimates of actual ATC accuracy, the differences in the mean estimates of acceptable ATC accuracy for the two groups were not statistically significant, $t\,(13) = 1.5$, $p > .05$.

Finally, the fourth question asked participants to provide any additional comments about the ATC, the imagery, or the study in general. Operators' responses to this question can be found in the Appendix. In general, their comments indicated that the limited number of scenes and vehicle configurations simplified their task of target localization. A few operators reiterated the point that the ATC must be reliable, or else it will be ignored, or will actually detract from the target acquisition task.

SECTION 4. DISCUSSION

The purpose of the present study was to quantify the effects of ATC cueing on target localization performance with first-generation FLIR imagery. The independent variables of interest, as mentioned earlier, included cueing (aided and unaided), ATC accuracy (50% and 75%), clutter site (open, treeline, and sparse), and range (4 km, 6 km, and 8 km). Our primary expectation, that the ATC would enhance performance accuracy relative to unaided performance, was not supported by the results. There were no differences in either the percentage of correct localizations or perceptual sensitivity between the aided and unaided conditions. This outcome is attributable in large part to unexpected ceiling effects. Specifically, performance in the unaided condition was already at a very high level (99.3%), leaving virtually no room for improvement when the ATC cues were available. It is our belief that the ceiling effects stem from the fact that the clarity of the digital imagery used in this study made it relatively easy to discern each vehicle, particularly since the ranges examined were in such close proximity to the target array.

Although cueing did not affect performance accuracy *per se*, it did affect operators' decision-making time and their confidence. Rather than enhancing speed, however, in the aided condition the cueing actually resulted in slower target localization times. With respect to confidence, on the other hand, subjective ratings indicated that operators were more confident in their decisions when the ATC cues were provided as compared to the unaided condition. Even though these results were statistically significant, it should be noted that the differences between the aided and unaided conditions were quite small. For example, operators were, on average, slower by only two-tenths of a second in the aided condition relative to the unaided condition. Similarly, on a scale that ranged from 1 to 6, confidence ratings in the aided condition were higher by only three-tenths of a point.

A Comparison of Present and Previous Study Results

The pattern of results in the present study is similar to the trends that emerged in an earlier investigation of the effects of cueing on target acquisition performance using medium resolution SAR imagery collected during the TESSA program (See, Davis, & Kuperman, 1997). In the SAR study, each patch map in the aided condition was presented, overlaid with four cue boxes representing the ATC's four highest regions of interest. Overall, cueing affected confidence, but not the percentage of correct localizations, perceptual sensitivity, or reaction time for correct localizations. Operators were more

confident in the aided condition ($M = 4.4$) than in the unaided ($M = 4.1$). A more in-depth analysis of ATC reliability in the SAR study revealed that the cueing did impact variables other than confidence, depending upon what objects in the scene were cued. Chiefly, when all of the ATC's cues were false alarms, both localizations and RT were significantly *worse* than if no aiding had been present at all. Conversely, when all three vehicles in the target array (TEL, MAN, and ZiL) were cued, observers' localization responses and perceptual sensitivity as well as their confidence in their decision-making were significantly better than in both the unaided condition and the aided condition where all cues were false alarms.

Reaction Time

In the SAR study, only when all of the ATC's cues were false alarms did the effect of cueing on RT slow decision-making time. Otherwise, aided RT did not differ from unaided RT. Conversely, in the present study, RT was slower overall in the aided condition as compared to the unaided condition. This outcome is most likely due to the manner in which the ATC information was displayed to operators. A dynamic presentation of FLIR imagery representing the aircraft's approach to the target array appeared first, followed by the presentation of the ATC's cues on a static frame that remained on the monitor for eight seconds. As several of the operators commented in their post-experimental questionnaires, they generally decided where they thought the target was located during the dynamic presentation. Even though they had reached a decision by the time the cues appeared, they still felt obligated to glance at them—a process that increased their reaction time as compared to the unaided condition. We should point out that the cues were presented in the manner just described because of the nature of the ATC's performance during the in-flight approach to the target array. Specifically, during the approach, the ATC's cues often jumped from object to object and did not remain fixed on a single item. Hence, to avoid this confusion in our study, we presented the ATC's cues for only the last frame in each range bin (i.e., at the end of the approach).

ATC Accuracy

A somewhat surprising outcome in the present study was the absence of an effect for ATC accuracy. Previous studies in the literature (Adams, 1991; Becker, Hayes, & Gorman, 1991; Entin & Entin, 1997; Entin, Entin, & Serfaty, 1996; Fulkerson, 1980; Jauer, Quinn, Hockenberger, & Eggleston, 1986; Kibbe & Weisgerber, 1991; Weisgerber & Savage, 1990) have indicated that ATCs must achieve a hit rate of 70% or higher with no more than four false alarms per image presentation if they are to enhance performance. Our results indicated no differences whatsoever between the low (50%) and high

(75%) accuracy groups. This outcome is most likely due to the ceiling effects mentioned earlier. Performance was already high in the unaided condition; hence, the cueing, regardless of whether it was high or low accuracy, could have little additional impact on performance.

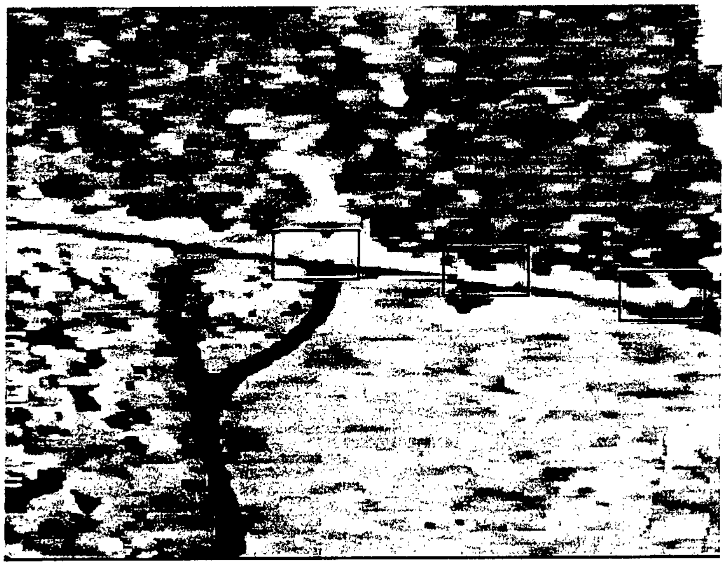
The ATC accuracy manipulation not only did not affect performance accuracy or confidence, but also did not affect participants' subjective estimates of the level of ATC accuracy experienced. Thus, regardless of the level of ATC accuracy actually presented, operators thought the level was about 50%. This outcome may be due to several different factors. First, as many participants stated, they often gave only a cursory glance at the cue boxes after having decided where they thought the TEL was positioned. Thus, because they were not attending to the cues, they may not have noticed how often the cues were accurate. Second, every image contained at least one ATC false alarm. Hence, while the ATC may have achieved a hit in some images, it always had at least one false alarm in every image. Some participants pointed this out in the questionnaire, stating that the accuracy for a given trial was at best 50% (one hit, one FA). Typically, ATC accuracy refers to the system's hit rate. Our designations of 50% and 75% accuracy levels thus referred to hit rates across the set of images presented and not to the false alarms. In responding to the questionnaire, some participants may have misinterpreted what we meant by accuracy. It may also be the case that operators take both hits and false alarms into account when estimating accuracy, and the occurrence of false alarms lowers their estimate of accuracy.

Interestingly, both groups of observers thought an ATC should be about 85% accurate to be useful, an estimate that exceeds the 70% level of accuracy that has been shown to be effective in previous studies (Adams, 1991; Entin & Entin, 1997; Entin, Entin, & Serfaty, 1996; Fulkerson, 1980; Jauer, Quinn, Hockenberger, & Eggleston, 1986; Kibbe & Weisgerber, 1991; Weisgerber & Savage, 1990). Although ATC-assisted performance in these studies was superior to unaided performance when the ATC was at least 70% accurate, it should be noted that the greatest performance differences occurred when the ATC achieved a 90% level of reliability (Kibbe & Weisgerber, 1991; Weisgerber & Savage, 1990). Further, although participants in the Becker, Hayes, and Gorman (1991) study rated all of the ATRs they examined as having at least some tactical advantage, none of the ratings exceeded a value of 60 on a scale of 100. The device that received the highest tactical value rating achieved a hit rate of .90 with only 0 or 1 false alarms per image. Thus, it may be the case that while ATCs with at least a 70% level of reliability can improve performance or confidence, operators would prefer to work with systems that achieve much greater levels of accuracy (as high as 85% or 90%).

Image Quality

Finally, in comparing the results from the unaided portion of the present study with those from a previous study of unaided detection/recognition using the same first generation TESSA FLIR imagery (See, Riegler, Fitzhugh, & Kuperman, 1996), we observed that performance was considerably better in the current study. For example, the overall mean percentage of correct responses was 93% (ranging from 70% to 99%) in the 1996 study as compared to 99% (ranging from 97.5% to 100%) in the present investigation. In the 1996 study, the FLIR imagery was received in analog format and was presented on a display similar in size to that used in the F-15E LANTIRN system. In the present study, the imagery was received in digital format and was presented on a Silicon Graphics O^2 computer monitor. We ensured that the visual angle from the image to the observer's eye was 10° in both studies; hence, the observed performance differences are not due to differences in visual angle. Instead, we believe they were due to: 1) differences in the appearance of the analog and digital imagery, as displayed on each monitor; and 2) the relatively small number of unique images available for presentation in the current study. With respect to image format, we first noted subjectively that the same analog images used in the 1996 study appeared much sharper and clearer when displayed as digital imagery on the O^2 monitor. These differences can best be seen by comparing the two panels in Figure 14 on the following page, which portrays the same treeline scene as it appeared in each study. As can be seen in the figure, the image from the current study was less fuzzy, making it possible for the operator to discern more detail and to better differentiate the vehicles from one another than in the 1996 study. Bear in mind that the task of localizing the TEL was enhanced in both studies by presenting multiple "still" frames in sequence, as opposed to presenting a single "still" frame, as shown here.

Second, the number of unique useable images was smaller in the present study ($N = 56$) as compared to the 1996 study ($N = 89$); at the same time, the total number of image presentations was larger in the present study ($N = 360$) than in the 1996 study ($N = 240$). This reduction in the size of the image set was due to the fact that we were combining an already limited set of useable imagery with information from the ATC, some of which also was not useable. The end result was a further reduction in the amount of useable imagery for the current study. Thus, operators saw many more repetitions of each image in the present investigation. As many of them commented in the post-session questionnaire, they knew where the target was located in each scene after only a few image presentations and did not need to search for it anew on subsequent presentations. This factor, in addition to the resolution differences between the two displays, may account for the improved performance in the present study.



Analog Imagery (1996 study)

Digital Imagery (current study)

Figure 14. Comparison of a treeline scene as it appeared in the 1996 and current studies.

SECTION 5. CONCLUSIONS

1. ATC-assisted target localization performance and perceptual sensitivity with first generation FLIR imagery did not differ from unaided performance accuracy and *d'*.

2. Operators were more confident in their decision-making when they were assisted by the ATC.

3. The reaction time for correct localizations was somewhat slower in the aided condition than in the unaided. Although this difference was small in magnitude, it indicates a need to present ATC information in a format that is useful but does not impede reaction time.

4. The manipulation of ATC accuracy (defined as an overall ATC hit rate of 50% or 75%) did not affect target localization accuracy, *d'*, RT for correct localizations, or confidence. Further, observers reported that the ATC's accuracy seemed to be about 50%, regardless of the actual level of accuracy presented. These effects were most likely due to the high level of performance in the present investigation.

5. The high level of performance at all of the range bin distances included in the present study suggests a need to explore ATC-assisted performance at ranges greater than 8 km.

REFERENCES

Adams, S. R. (1991). *Presentation format as a determinant of targeting task performance* (Report No. NWC TP 7125). China Lake, CA: Naval Weapons Center.

Aerospace Daily (April 23, 1996). *ATR system R & D should stay at low level.* Volume 178 (17), 137 (1).

Augustyn, K. (1992). A new approach to automatic target recognition. *IEEE Transactions on Aerospace and Electronic Systems, 28 (1),* 105-114.

Becker, C. A., Hayes, B. C., & Gorman, P. C. (1991). *User acceptance of intelligent avionics: A study of automatic-aided target recognition* (Report No. NASA CR-177583). Moffett Field, CA: Ames Research Center.

Beideman, L. R., Gomer, F. E., & Levine, S. H. (1980). *Dynamic target acquisition: Empirical models of operator performance* (Report No. AFOSR-TR-80-1177). Bolling Air Force Base, DC: Life Sciences Directorate, Air Force Office of Scientific Research. (DTIC Report No. AD A092263)

Carr, K. T. (1988). An investigation into the effects of a simulated thermal cueing aid upon air-to-ground search performance. In D. Brogan (Ed.), *Visual Search* (pp. 361-370). New York: Taylor & Francis.

Craig, A. (1984). Human engineering: The control of vigilance. In J. S. Warm (Ed.), *Sustained attention in human performance* (pp. 247-291). Chichester, UK: Wiley.

Davis, I.E., See, J. E., Shacklett, S. L. & Kuperman, G. G. (1996). *Unaided target acquisition performance with enhanced APG-70 Synthetic Aperture Radar: A signal detection theory analysis* (Report No. AL/CF-TR-1996-0104). Wright-Patterson Air Force Base, OH: Armstrong Laboratory.

Delashmit, W. H. (1989). Present status and future needs for automatic target recognizers. *SPIE Digital Image Processing Applications, 1075,* 373-383.

Entin, E. B., & Entin, E. E. (1997). The effects of decision time and decision-aid accuracy on workload and performance. In *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting*, (pp. 177-181). Santa Monica, CA: The Human Factors Society.

Entin, E. B., Entin, E. E., & Serfaty, D. (1996). Optimizing aided target-recognition performance. In *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting*, (pp. 233-237). Santa Monica, CA: The Human Factors Society.

Fulghum, D. A. (1994). Scud hunting may drop under 10 minute-mark. *Aviation Week and Space Technology, 140 (8), 90.*

Fulkerson, D. C. (1980). Required performance levels for autocueing devices. In *Proceedings of the Human Factors Society 24th Annual Meeting*, (pp. 511-515). Santa Monica, CA: The Human Factors Society.

Gescheider, G. A. (1985). *Psychophysics: Method, theory, and application* (2nd ed.). Hillsdale, NJ: Erlbaum.

Goble, L. G., Williams, L. G., Pratt, P. D., Wald, J., Rubin, G., & Hanson, T. (1980). *Human operator interface with FLIR displays* (Report No. AMRL-TR-79-114). Wright-Patterson Air Force Base, OH: Aerospace Medical Research Laboratory.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Hacker, M. J., & Ratcliff, R. (1979). A revised table of d′ for M-alternative forced choice. *Perception & Psychophysics, 26*, 168-170.

Huynh, H., & Feldt, L. S. (1970). Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. *Journal of the American Statistical Association, 65*, 1582-1589.

Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics, 1*, 69-82.

Jauer, R. A., Quinn, T. J., Hockenberger, R. L., & Eggleston, R. G. (1986). *Radar aided mission/aircraft capability exploration (RAM/ACE): Full task simulation* (Report No. AAMRL-TR-86-015). Wright-Patterson Air Force Base, OH: Armstrong Laboratory.

Jones, D. T. (1997). *Combat air forces concept of operations for command and control against time critical targets.* (OPR: HQ ACC/DRAW) Langley Air Force Base, VA.

Kibbe, M. P., Adams, S., Weisgerber S. & Savage, S. (1990). *Aircrew opinions on targeting autoclassifiers* (Report No. NWC TP 7057). China Lake, CA: Naval Weapons Center.

Kibbe, M. P., & Weisgerber S. (1991). *Operator use of automated target recognition (ATR) systems under time constrained conditions* (Report No. NWC TP 7136). China Lake, CA: Naval Weapons Center.

Kuperman, G. G. (1997). Human system interface (HSI) issues in assisted target recognition (ASTR). *Proceedings of the 1997 IEEE National Aerospace and Electronics Conference,* (pp. 37-48). Piscataway, NJ: IEEE.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* Cambridge: Cambridge University Press.

Mumma, D. R., & Bell, H. H. (1993). *F-15E Critical mobile target mission simulation data* (Report No. AL-TP-1992-0064). Brooks Air Force Base, TX: Armstrong Laboratory.

Pryce, J. D. (1995). *TESSA SAR/FLIR data collection report. Volume 1* (Report No. OMI-560). Dayton, OH: Veridian/Veda Operations.

Rotman, S. R., Kowalczyk, M. L., Cartier, J., & Chang, J. (1994). Understanding infrared clutter through fixation point analysis. *Infrared Imaging Systems: Design, Analysis, Modeling and Testing V: Proceedings of the International Society for Optical Engineering (SPIE) Vol. 2224,* 142-154.

See, J. E., Davis, I., & Kuperman, G. G. (1997). *Automatic target cueing and operator performance with enhanced APG-70 synthetic aperture radar imagery* (Report No. AL/CF-TR-1997-0171). Wright-Patterson Air Force Base, OH: Armstrong Laboratory.

See, J. E., & Kuperman, G. G. (1995). *The effects of resolution and other factors on target acquisition performance using simulated synthetic aperture radar imagery* (Report No. AL/CF-TR-1995-0149). Wright-Patterson Air Force Base, OH: Armstrong Laboratory.

See, J. E., Riegler, J. T., Fitzhugh, E., & Kuperman, G. G. (1996). *Unaided target acquisition performance with first-generation forward looking infrared imagery: A signal detection theory analysis* (Report No. AL/CF-TR-1996-0094). Wright-Patterson Air Force Base, OH: Armstrong Laboratory.

See, J. E., Warm, J. S., Dember, W. N., & Howe, S. R. (1997). Vigilance and signal detection theory: An empirical evaluation of five measures of response bias. *Human Factors, 39*, 14-29.

Shumaker, D. (1979). Thermal-imaging-system (TIS) dynamic field performance. In F. A. Rosell & G. Harvey (Eds.), *The fundamentals of thermal imaging systems* (Report No. NRL 8311, pp. 111-142). Washington, D.C.: Naval Research Laboratory. (DTIC Report No. AD A073763)

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*, 34-50.

Strzempko, T., & Pritchard, D. (1990). Designer's handbook: Surveillance applications of infrared imaging. *Photonics Spectra, 24(4),* 127-132.

Toms, M. L., & Kuperman, G. G. (1991). *Sensor fusion: A human factors perspective* (Report No. AL-TR-91-0152). Wright-Patterson Air Force Base, OH: Armstrong Laboratory.

Turner, S. (1995). *Quantifying human performance of a dynamic military target detection task: An application of the theory of signal detection.* Unpublished master's thesis, Wright State University, Dayton, OH.

Valeton, J. M., & Bijl, P. (1994). *Target acquisition: Human observer performance studies and TARGAC model validation* (Report No. ERO R/D-6720-EN-01). Soesterberg, Netherlands: TNO Human Factors Research Institute, U. S. Army European Research Office. (DTIC Report No. AD A284436)

Weisgerber, S. A., & Savage, S. F. (1990). *Operator ship classification using an automatic target recognition (ATR) system in conjunction with forward looking infrared (FLIR) imagery* (Report No. NWC TP 7101). China Lake, CA: Naval Weapons Center.

Wilson, D. L. (1992). *Theory of signal detection and its application to visual target acquisition: A review of the literature* (Report No. AL-TR-1992-0083). Wright-Patterson Air Force Base, OH: Armstrong Laboratory.

GLOSSARY

AO	Attack Operations
ASC/FBXT	U.S. Air Force Aeronautical Systems Center/Theater Missile Defense Integrated Product Team
ATC	Automatic Target Cueing
ATR	Automatic Target Recognition
c	Response Bias Index
C^2	Command and Control
CIWAL	Crew-Aiding and Information Warfare Analysis Laboratory
CONOPS	Concept of Operations
DoD	Department of Defense
d'	Perceptual sensitivity
FA	False Alarm
FLIR	Forward-Looking Infrared
H	Hit
HIPAC	High Pressure Air Compressor
HMI	Human Machine Interface
km	Kilometers
KTGS	Knots True Ground Speed
LANTIRN	Low Altitude Navigation and Targeting Infrared for Night
M	Mean
nmi	Nautical Mile
PAC - 2	Patriot Advanced Capabilities
RT	Reaction Time
SAR	Synthetic Aperture Radar

SD	Standard Deviation
SG	Silicon Graphics
TAD	Theater Air Defense
TCT	Time Critical Target
TEL	Transporter/Erector/Launcher
TESSA	TMD Eagle Smart Sensor and ATC
TM	Theater Missile
TMD	Theater Missile Defense
TSD	Theory of Signal Detection
z	Inverse Normal Distribution

APPENDIX A: QUESTIONNAIRE RESPONSES

The following is a summary of the operators' responses on the questionnaire administered upon completion of data collection. The following abbreviations have been used to identify the aircrew experience of the respondents: Pilot (P), Instructor Pilot (IP), Instructor Radar Navigator (IRN), Instructor Navigator (IN), and Standards Evaluator (STAN/EVAL). Note that some respondents were experienced in more than one category. Any material in brackets corresponds to the experimenter's annotations.

Question #1. "Did the ATR help with your targeting decision ?" A tally across the 16 observers yielded an equal number of yes and no responses (8 each). Of note, was the observation that each accuracy group was equally divided within the two responses.

Comments Among the 'Yes' Respondents:

<u>50% Accuracy Group</u>

"When it was on [the boxes], it reaffirmed the target (didn't need to look all over terrain), if it was in error, then it still usually focused attention to [the] right area." (S#1: IP)

"It helped confirm my target choice. When the box didn't come up where I had target[ed], did quick search but usually went with my original choice." (S#9: IRN, STAN/EVAL)

"Not always [helpful], but sometimes it helped focus to correct target area. ([ATC] narrowed [the] search pattern)." (S#13: IP, STAN/EVAL)

<u>75% Accuracy Group</u>

"[The ATR] made me more confident. Even if it designated something else." (S#14: IRN)

"If it agreed with my decision then it would increase my confidence, but if it did not agree with me I ignored it." (S#18)

"Faster decision time by confirming locus of targets." (S#16: IN, WSO)

"ATR provided a reasonably assured 'possibility' that was given first consideration. Although not always true [accurate], cases where the box was misplaced proved deceptive about 30% [of the time when the] area closely looked like the target." (S#8: IP, STAN/EVAL)

Comments Among the 'No' Respondents:

50% Accuracy Group

"By the time the ATR box appeared I had already spotted the target and would give a courtesy glance at the box to see why it was so far off. ATR never found a target I hadn't already seen." (S#5: P, IP)

"ATR seemed to land on objects other than the SCUD most of the time." (S#11: IN)

"More often than not, it was wrong...breeds no confidence." (S#7: P)

"NEGATIVE ! With the poor accuracy of this system, it does nothing but question your judgment and distract you, thereby decreasing your confidence level and increasing decision time. Confidence and quick decision making are very important in a tactical environment." (S#15: WSO)

75% Accuracy Group

"The target box wasn't a deciding factor." (S#12)

"I did not change my target choice at all due to the boxes. A couple of times I had difficulty selecting a target, but the boxes did not help. Whenever the boxes did not agree with my choice, I checked both boxes, but never selected one of them." (S#4: IN, STAN/EVAL)

Too many extraneous target boxes. I usually had my mind made up before the boxes came up. Used more in confirmation if I used it at all." (S#10)

"Tried to focus on target, not ATR. Plus I found its accuracy questionable on several occasions." (S#6)

Question #2. "What is your estimate of ATR accuracy ?" The 50% Accuracy Group yielded an average estimate of 48.125% (*SD* =15.104%). These estimates ranged from 20% to 70%, with half of the observers correctly reporting 50%. The 75% Accuracy Group yielded an average estimate of 50.625% (*SD* =20.777%). These estimates ranged from 10% to 70%, with no observers correctly reporting 75%. It should be mentioned that a couple of participants stated that placement accuracies of

the ATC boxes were never greater than 50% for a given trial (1 box was always overlaid on an area other than the TEL), a valid interpretation that may have affected their estimates.

Question #3. "What would you consider an 'acceptable' ATR accuracy ?" The 50% accuracy group responded with 83.571% (*SD* =7.480%), with a range from 70% to 95%. The 75% accuracy group reported 89.375% (*SD* =7.289%), with a range from 75% to 95%. The two groups overall, yield a value of 86.667% (*SD* =7.715%).

Question #4. Other comments:

50% Accuracy Group

"Because of the limited number of scenes (3 as best I could tell) the surrounding terrains and features (roads, intersections, tire tracks) lead me to the low contrast targets (familiarity). Then it was simply giving a confidence level that the selected target could not have been something else." (S#5: P, IP)

"Easier to determine target when all three vehicles could be seen in field of view." (S#13: IP, STAN/EVAL)

"If accuracy [of the ATC] is not too good then would tend to ignore." (S#1: IP)

"[The ATC is a] great help to confirm operator 'RSI', but could detract if not high correlation with placement on correct target." (S#9: IRN, STAN/EVAL)

"The [ATC] accuracy in this experiment was completely unacceptable; as an [WSO] instructor, I would advise my students to keep the system turned off unless at a point where they had to release ordnance and had zero confidence in their own designation at that time. The system is of no value unless it is at least as accurate as I am. Otherwise it only slows down the targeting process because I have more votes to count and assess."

[On the target video used in this experiment] "These were much too easy on the average. There have been many times I've come back from a mission without having found my target, even when it was a large bridge. At long range, in poor IR conditions, or when many target distracters are present is when the WSO needs help. But it must be accurate help. This experiment would be more realistic if there were more distracters, such as buildings with similar size and shape to the target. Or put the target in the trees

on a narrow road, etc. There were a few too many runs, making it hard to keep an unbiased "first look" judgment."

[On the subject of the test participants selected for this experiment] "I feel using any test subjects other than genuine users of Nav FLIR/Target FLIR systems would be a detriment to the experiment. Whether the ATR helps or hurts the accuracy/confidence level of the operator is almost totally dependent on his tactical experience. There are many other important factors that affect a targeting decision, such as the written Rules of Engagement (ROE), allowable collateral damage, threat of air-to-air engagements, threat of SAM/AAA in the target area (I may not want to get any closer to the target to increase my designation confidence if the threat is high). This experiment would only be meaningful if taken to an F-15E squadron (or several) and getting a large sample size from people that know what target selection in a tactical environment is like. I fear anyone else (even flyers) would adversely skew the data and risk fielding systems that the users dislike." (S#15: WSO)

"At longer ranges it [ATR] will probably help pick up the target. During the test [experiment], many times I could make out the missile on the TEL - obviously that was the target. If the ATR (or test) could simulate a greater distance from the target and still recognize the target that would aid in narrowing the field of search and confirming the 'designated' target was correct." (S#11: IN)

75% Accuracy Group

"Bad info is worse than no info. The dynamic scenes, length to width ratios, [and] shadows made a lot more difference than target boxes." (S#10)

"The ATR had extreme trouble on edge-wise to the treeline (up to 45° of angle). It [ATR] never selected the target." (S#4: IN, STAN/EVAL)

"After seeing how poor the ATR accuracy was, I didn't trust it enough to even consider it in my decisions." (S#12)

"Remove the # of runs [trials] from the display - 'rush' or 'push' to go forward fueled by feedback of progress toward the total count." (S#8: IP, STAN/EVAL)

"2 uses for ATR: 1) Identify target area (then declutter) 2) Confirm selection of target." (S#16: IN, WSO)

"This was a benign environment for the test. Throw in other vehicles, and I think it accurately would decline." (S#14: IRN)